

**VALIDATING DIFFUSION SPECTRUM IMAGING-BASED FIBER TRACTOGRAPHY
FOR COGNITIVE NEUROSCIENCE RESEARCH**

by

Jeffrey S. Phillips

Bachelor of Arts, Villanova University, 1997

Master of Science, University of Pittsburgh, 2008

Submitted to the Graduate Faculty of
the Kenneth P. Dietrich School of
Arts and Sciences in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH
THE KENNETH P. DIETRICH
SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Jeffrey S. Phillips

It was defended on

May 1, 2012

and approved by

Dr. Julie Fiez, Professor, Department of Psychology

Dr. Christian Schunn, Professor, Department of Psychology

Dr. Robert Kass, Professor, Department of Statistics, Carnegie Mellon University

Dissertation Advisor: Dr. Walter Schneider, Professor, Department of Psychology

Copyright © by Jeffrey S. Phillips

2012

VALIDATING DIFFUSION SPECTRUM IMAGING-BASED FIBER TRACTOGRAPHY FOR COGNITIVE NEUROSCIENCE RESEARCH

Jeffrey S. Phillips, PhD.

University of Pittsburgh, 2012

White matter fiber tractography based on diffusion-weighted magnetic resonance imaging is a promising method for non-invasive investigation of anatomical connectivity in the human brain. Knowledge of the white matter connections linking functional brain areas can inform interpretation of functional imaging results and allow the construction of biologically informed computational and statistical models. However, relatively little attention has been paid to the reproducibility and external validity of tractography results, even as the user base of this technology continues to grow, and as tractography research is applied to cognitive neuroscience research in novel ways. In this investigation, we addressed the reliability and validity of deterministic tractography results based on diffusion spectrum imaging (DSI). Reliability was evaluated both in terms of the presence/absence of fiber connections across sessions and the correlation of fiber density values. Validity was assessed by comparing tractography results to findings from invasive studies of the macaque monkey: we focused on the cortical and subcortical connections of the frontal eye fields (FEF). Results indicated significant variability in tractography: on average, intercortical connections present in one session had only a 75% likelihood of being detected in a second session from the same individual. However, the fiber density of repeatedly-detected connections was highly reliable, with an average between-session

correlation coefficient of 0.94. Next, we investigated how global vs. targeted tractography approaches affected reliability and detection power. We found that a targeted approach, involving the use of region-of interest (ROI) constraints, yielded a large advantage in detection power and modest improvements in reliability. Finally, fiber connections of the human FEF were broadly consistent with hypotheses derived from a meta-analysis of macaque findings: we found reliable projections to the supplementary eye fields (SEF), striatum, thalamus, and parietal cortex. In contrast, we found lesser connectivity to a set of foil regions. The combined results of this study validate the use of DSI-based fiber tractography to address hypotheses relating to human brain connectivity. However, widespread noise in tractography results highlights the need for conservative approaches to fiber tracking research. We especially emphasize the benefits of collecting multiple data samples per participant and of addressing targeted hypotheses.

TABLE OF CONTENTS

PREFACE.....	xiii
1.0 EXPERIMENT 1: BETWEEN-SESSION RELIABILITY OF DIFFUSION SPECTRUM IMAGING-BASED FIBER TRACTOGRAPHY	1
1.1 INTRODUCTION	1
1.2 METHOD	6
1.2.1 Participant information.....	6
1.2.2 Procedure.....	7
1.2.3 Anatomical Image Acquisition and Processing.....	7
1.2.4 Diffusion Image Acquisition and Processing.....	8
1.2.5 Anatomical Region-of-Interest Creation.....	11
1.2.6 Deterministic Tractography Methods.....	12
1.2.7 Intercortical Connectivity Reliability Analysis.....	14
1.2.8 Targeted Tract Reliability Analysis.....	18
1.2.9 Targeted vs. Global Tracking Results: Region-to-Region Analysis.....	21
1.2.10 Targeted vs. Global Tracking Results: Voxelwise Analysis.....	22
1.2.11 Identifying Sources of Between-Session Error in Tractography Results.....	24
1.2.11.1 Assessment of Reseeding Error.....	27

1.2.11.2	Parameter Sensitivity Analysis.....	28
1.2.11.3	Assessment of Spatial Error.....	31
1.2.11.4	Assessment of Acquisition-Related Error.....	32
1.3	RESULTS.....	34
1.3.1	Intercortical Connectivity Reliability Analysis.....	34
1.3.1.1	Between-Session Correlation of Connection Density.....	34
1.3.1.2	Reliability Across Individuals and Sessions.....	41
1.3.2	Targeted Tractography Results.....	56
1.3.3	Global Tractography Results.....	61
1.3.4	Targeted vs. Global Tracking Comparison: Region-to-Region Connectivity.....	63
1.3.5	Targeted vs. Global Tracking Comparison: Voxelwise Analysis.....	69
1.3.6	Sources of Error in Tractography: Summary of Follow-up Analyses.....	77
1.3.6.1	Assessment of Reseeding Error.....	77
1.3.6.2	Parameter Sensitivity Analysis.....	79
1.3.6.3	Assessment of Spatial Error.....	87
1.3.6.4	Assessment of Acquisition-Related Error.....	92
1.4	DISCUSSION.....	96
1.4.1	Between-Session and Between-Subject Reliability.....	96
1.4.2	Most-Often-Detected Intercortical Connections.....	101
1.4.3	Global vs. Targeted Tracking Comparisons.....	104
1.4.4	Error Source Analyses.....	107
1.5	INTERIM CONCLUSION.....	110

2.0	EXPERIMENT 2: VALIDATING DSI-BASED TRACTOGRAPHY AGAINST KNOWN ANATOMICAL CONNECTIONS OF THE FRONTAL EYE FIELDS.....	112
2.1	INTRODUCTION	112
2.2	METHOD	115
2.2.1	Participant Information.....	115
2.2.2	FMRI Task Design and Procedure.....	116
2.2.3	Image Acquisition and Processing.....	117
2.2.4	Deterministic Tractography Methods.....	119
2.2.5	Macaque Meta-Analysis and Human ROI Definition.....	121
2.2.6	Statistical Methods and Hypotheses.....	129
2.3	RESULTS.....	130
2.3.1	ROI Definition.....	130
2.3.2	Target vs. Foil Comparison.....	133
2.3.3	Topography of Target Projection Endpoints in FEF.....	145
2.3.4	Effects of Anisotropy Threshold.....	146
2.4	DISCUSSION.....	149
2.5	INTERIM CONCLUSION.....	151
3.0	GENERAL DISCUSSION.....	152
4.0	CONCLUSION.....	163
	APPENDIX.....	165
	BIBLIOGRAPHY	166

LIST OF TABLES

Table 1. Target and foil regions for parameter sensitivity test.....	30
Table 2. Within-subject, between-session reliability of intercortical connectivity matrices.....	35
Table 3. Left- and right-hemisphere areas with highest degree of MOD connections.....	50
Table 4. Brain areas of highest degree in long-distance subset of MOD connection network.....	55
Table 5. Left corticospinal tract: fiber counts from global tractography.....	61
Table 6. Left arcuate fasciculus: fiber counts from global tractography.....	62
Table 7. Callosal projection fibers: fiber counts from global tractography.....	62
Table 8. Difference between correlation coefficients from targeted vs. global tracking.....	67
Table 9. Descriptive statistics for endpoint maps from global vs. targeted tracking.....	70
Table 10. Descriptive statistics for full-tract density maps from global vs. targeted tracking.....	71
Table 11. Comparison of correlations between paired B0 images vs. paired QA0 maps.....	94
Table 12. Low, medium, and high QA thresholds for FEF tractography.....	120
Table 13. Summary of CoCoMac query of FEF connectivity.....	125
Table 14. Mask sizes for left and right FEF.....	132
Table 15. Significance of FEF connections to foil ROIs.....	139
Table 16. Significance of FEF connections to target ROIs.....	141
Table 17. Count of detected FEF connections to target ROIs.....	143
Table 18. Significance of target ROI connections relative to foil connection density.....	144

LIST OF FIGURES

Figure 1. Overview of Experiments 1 and 2.....	4
Figure 2. Intercortical connectivity analysis: Pearson’s correlation metric.....	15
Figure 3. Intercortical connectivity analysis: overlap metric.....	16
Figure 4. Between-session comparison of log-transformed intercortical connection densities.....	37
Figure 5. Detection frequency for intercortical connections.....	39
Figure 6. Scatterplot of density values for connections detected in all datasets.....	40
Figure 7. Relationship between connection density and frequency of detection.....	42
Figure 8. Distribution of connection densities by connection type.....	44
Figure 9. Histograms of connection density by frequency of detection.....	45
Figure 10. Type and symmetry of interhemispheric connections replicated in all datasets.....	47
Figure 11. Cortical areas in the network of MOD (most-often-detected) connections.....	49
Figure 12. Graph of symmetric intrahemispheric MOD connections.....	51
Figure 13. Graph of asymmetric intrahemispheric MOD connections.....	52
Figure 14. Graph of interhemispheric MOD connections.....	53
Figure 15. Distribution of average fiber lengths in MOD connection set.....	55
Figure 16. Targeted tracking of left hemisphere corticospinal tract.....	57
Figure 17. Targeted tracking of the left arcuate fasciculus.....	58
Figure 18. Targeted tracking of interhemispheric callosal projection fibers.....	59

Figure 19. Between-session overlap for global and targeted tracking.....	63
Figure 20. Between-session reliability of fiber density with global and targeted tracking.....	66
Figure 21. Comparison of detection power under global vs. targeted tracking conditions.....	69
Figure 22. Voxelwise between-session overlap for global and targeted tracking.....	73
Figure 23. Voxelwise between-session density correlations for global and targeted tracking.....	74
Figure 24. Voxelwise detection power for global and targeted tracking.....	75
Figure 25. Assessment of reseeded error.....	78
Figure 26. Results from Step 1 of the parameter sensitivity analysis.....	80
Figure 27. Results from Step 2 of the parameter sensitivity test.....	84
Figure 28. Top-scoring corticospinal tract results from parameter sensitivity test.....	85
Figure 29. Mutual similarity of tractography results from Step 2 of parameter test.....	86
Figure 30. Reliability scores for tract pairs before and after spatial alignment.....	90
Figure 31. Quality check of affine-based spatial transformation between sessions.....	91
Figure 32. Between-session comparison of quantitative anisotropy (QA) maps.....	94
Figure 33. Voxels masked out in one session but not the other.....	95
Figure 34. Oculomotor task used to localize the frontal eye fields.....	117
Figure 35. Map of target and foil FEF cortical projection areas.....	124
Figure 36. High-density target and low-density foil masks for FEF tractography.....	129
Figure 37. Combined mask sizes for high-density targets and low-density foils.....	133
Figure 38. Fiber connections from FEF to subcortical targets.....	135
Figure 39. Fiber connections from FEF to posterior parietal cortex.....	136
Figure 40. Fiber connections from FEF to the superior frontal gyrus.....	137
Figure 41. Mean fiber density of connections to foil ROIs.....	138

Figure 42. Mean fiber density of connections to target ROIs.....	140
Figure 43. Endpoint fields of subcortical projections to FEF.....	147
Figure 44. Endpoint fields of cortical projections to FEF.....	148

PREFACE

My heartfelt thanks go out to my advisor, Dr. Walter Schneider, who gave me the time and space to find a voice for myself in cognitive neuroscience, and who showed tremendous kindness toward me upon the death of my mother. I must likewise thank Dr. Julie Fiez, who has been a gracious mentor to me since I began my research career with her in 2003. For technical assistance with tractography methods in this dissertation, I owe much gratitude to Sudhir Pathak, a casual genius whom I hope to count as a friend and colleague in the years to come. I must also thank Frank Yeh for responding to questions and bug reports—he facilitated my research enormously in doing so. Emily Braun provided me with endless encouragement and made the Schlab a brighter place. Very personal thanks to my father and sister for their support—I know they are glad to see me earn the second doctorate in our family. Finally, I would like to express my love and appreciation for my wife, Nina Clements, who sustained me throughout graduate school with good counsel, wonderful food, and understanding.

1.0 EXPERIMENT 1: BETWEEN-SESSION RELIABILITY OF DIFFUSION SPECTRUM IMAGING-BASED FIBER TRACTOGRAPHY

1.1 INTRODUCTION

Increasingly, cognitive neuroscientists seek to test hypotheses that involve complex functional networks. The nodes of such functional networks must necessarily be joined by either direct or indirect anatomical connections, in the form of white matter fiber bundles. However, direct evidence of anatomical connectivity is usually unavailable, due to methodological limitations or the use of human participants, which precludes invasive study of brain connectivity. Rather, inferences about network structure are informed by statistical models of functional imaging data (Cohen et al., 2008; Cole et al., 2010; Dosenbach et al., 2007; Fox & Raichle, 2007; Rogers et al., 2007), behavioral observations following brain injury or experimental intervention (Nomura et al., 2010), or prior neuroanatomical evidence of anatomical connectivity (Modha & Singh, 2010). Most current knowledge of brain connectivity comes from invasive studies of non-human animals, using such techniques as autoradiographic tracer labeling (Huerta et al., 1986, 1987; Parthasarathy et al., 1992); some additional data comes from post-mortem dissection of human brains (Bürgel et al., 2005). However, these data are

costly to acquire, especially because tissue volumes or whole brains can often be used only for a single test. As a consequence, knowledge of brain connectivity depends largely on the synthesis of findings gathered piecemeal using different methods, from different species and individuals. The heterogeneity of data sources limits the precision and confidence of our knowledge about brain connectivity, particularly when researchers wish to make assertions or inferences about the human brain.

The combination of diffusion-weighted imaging (DWI) and fiber tractography offers a way to non-invasively investigate anatomical connectivity in the human brain. Techniques such as diffusion spectrum imaging (DSI; Wedeen et al., 2005, 2008) and high angular resolution diffusion imaging (HARDI) have the ability to reproduce known features of human brain connectivity with high levels of detail (Behrens et al., 2003; Greenberg et al., 2012; Jarbo et al., 2012; Phillips et al., 2012; Pyles et al., 2012; Verstynen et al., 2011), providing clinically useful information (Fernandez-Miranda et al., 2010). The strength of these methods lies in their high spatial and angular resolution, which allows the construction of rich diffusion representations using model-free techniques, like the orientation distribution function (Seunarine & Alexander, 2009). Such representations can be used to resolve complex fiber configurations (Wedeen et al., 2005) that have confounded simpler forms of DWI, such as diffusion tensor imaging (DTI; Basser et al., 1994a, b).

However, important practical and theoretical questions remain regarding the applicability of DWI-based tractography to cognitive neuroscience research. Chief among these questions are the reliability and validity of tractography results. Questions about reliability include the replicability over time of tractography results in a given individual, both in terms of spatial extent and connection density. Unknown factors about the validity of tractography results include the

incidence of false positives and false negatives, as well as their proportions to one another. Resolution of these questions is important for understanding the noise levels inherent in DSI data, as well as the specificity vs. generalizability of tractography results.

Two previous studies have investigated between-session reliability of DSI-based tractography (Bassett et al., 2011; Cammoun et al., 2012). These studies have reported high within-subject correlations of connection density values, suggesting that tractography results are highly reproducible over time. Moreover, both studies have analyzed reliability at different levels of spatial resolution, finding that the reproducibility of results decreased as the spatial scale of analysis grew finer. While these results are informative, these studies have left several questions unaddressed. Perhaps most importantly, neither study appears to report the reproducibility of connectivity results in the simplest terms possible: the likelihood that a connection which is present or absent in one dataset will have the same status in a repeat observation. Furthermore, investigation of possible explanations for between-session error was limited: Cammoun et al. (2012) addressed only one source of variability, the effect of randomness in ROI generation. In contrast, comprehensive investigations of possible error sources have been performed for tractography based on DTI (Lori et al., 2002) and for functional magnetic resonance imaging (e.g., Zou et al., 2005).

In the present study, we sought to replicate and extend the findings of Bassett et al. (2011) and Cammoun et al. (2012). Figure 1 gives an overview of the analyses included in Experiment 1. First, we reproduced Cammoun et al.'s test of reliability in intercortical connection matrices, based on regions of interest (ROIs) automatically defined by FreeSurfer (Desikan et al., 2006; Fischl et al., 2002, 2004). Based on prior results, we predicted that we would find high between-session correlation of fiber density values. Additionally, we examined the reproducibility of

binarized results, where connections were coded as being simply present or absent. This analysis yielded information that was complementary to the correlation of fiber density values for repeatedly-observed connections. Given the high correlations reported by Bassett et al. (2011) and Cammoun et al. (2012), we predicted that the reproducibility of binarized results would also be high.

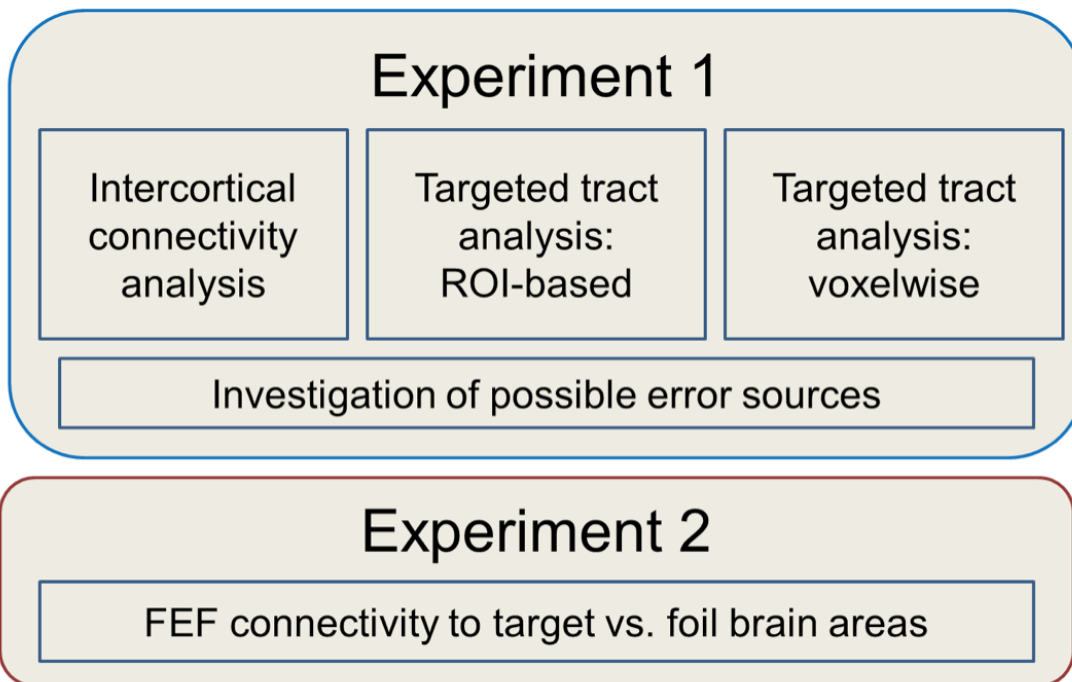


Figure 1. Overview of Experiments 1 and 2.

We were additionally interested in procedural effects upon tractography results and their reliability. The competitive context for seed generation and fiber propagation represented one

such variable of interest. Tractography investigations commonly use a global seeding and tracking approach, in which fibers are propagated from starting points throughout the entire brain, or large portions of it, with minimal constraints on fiber trajectory or target. We refer to this approach as global tractography; other authors have referred to it by the term “exhaustive search” (Bassett et al., 2011; Gong et al., 2008). In contrast, deterministic tractography can also be performed with specific ROI constraints, and with tracking parameters optimized for detecting expected fibers of interest. We refer to this approach as targeted tractography; our research group has previously used a targeted approach for tracking fibers of the corticospinal tract (Verstynen et al., 2011), low-detectability callosal projections to parietal and posterior temporal cortex (Jarbo et al., 2012), and attendotopic projections from early visual regions to parietal cortex (Greenberg et al., 2012; Phillips et al., 2012).

We predicted that targeted tracking would be associated with superior between-session reliability and detection power, compared to global tractography. We reasoned that targeted tracking could improve reliability in several ways. First, targeting specific fiber tracts allows the imposition of more ROI-based constraints than in a whole-brain or nearly whole-brain tracking job. Such constraints can counteract “diffusion steal effects”, in which a deterministic tractography algorithm may consistently fail to follow a valid diffusion pathway in favor of a stronger, adjacent pathway (for example, see Jarbo et al., 2012). Second, and relatedly, targeted tracking can be computationally more efficient for detecting specific fibers of interest, allowing fuller saturation of possible diffusion pathways than global tracking. Finally, targeted tracking allows the user to optimize tractography parameters to increase detection power, whereas global tracking jobs must use a single set of parameters to capture fiber tracts with very different morphology.

To investigate the effects of global vs. targeted tracking, we performed more detailed analysis on three specific tracts of interest: the arcuate fasciculus (AF), the corticospinal tract, and the interhemispheric projection fibers of the corpus callosum. For these tracts, we performed both ROI-based and voxelwise analyses. As in the intercortical connectivity analysis, ROI-based analyses indicated the reproducibility of region-to-region connections. Additionally, voxelwise analysis of tract endpoint maps was performed to assess the reproducibility of termination fields, which are of special interest for their comparability to functional imaging data. Finally, we performed voxelwise analysis of full-tract density maps as a means of comparing the morphology of tract bodies across sessions. In both ROI-based and voxelwise analyses, we additionally examined the effect of tracking condition (global vs. targeted) on detection power, assessed by the absolute number of connections and/or voxels reliably detected across sessions.

1.2 METHOD

1.2.1 Participant Information

Seven healthy adults between the ages of 22 and 34 participated. The mean age was 26.7 (4.5) years. Four participants were male, and three were female. All participants gave informed consent as required by the Institutional Review Board of the University of Pittsburgh. Five people received monetary compensation for their participation, and two co-authors volunteered to participate.

1.2.2 Procedure

Four participants underwent 2 two-hour scanning sessions each for the 2009 Pittsburgh Brain Connectivity Competition. In a total of four of these sessions, high angular resolution diffusion imaging (HARDI) data were also collected. One participant, for whom HARDI data were not collected, underwent functional magnetic resonance imaging (fMRI) in both sessions. Data for the remaining three participants were collected for an unpublished cognitive training experiment; scanning sessions for these participants were either 90 minutes or two hours and included both DSI and fMRI data collection. DSI and tractography results for all seven participants have been reported in previous publications by our group (Jarbo et al., 2011; Verstynen et al., 2011).

1.2.3 Anatomical Image Acquisition and Processing

High-resolution T1-weighted anatomical images were collected for all subjects using a magnetization-prepared rapid gradient echo (MPRAGE) sequence. (TR=2110 ms, TE=2.63 ms, flip angle=8°, 176 slices, field-of-view=256 x 256 mm, voxel size = 1.0 x 1.0 x 1.0 mm). For a subset of datasets, T1 images were collected with voxel sizes of 0.5 x 0.5 x 0.5 mm; prior to other processing steps, these images were resampled to match the 1 mm isotropic voxel size of the other T1 images.

Anatomical images were processed using the FreeSurfer recon-all pipeline (Desikan et al., 2006; Fischl et al., 2002, 2004). Input images were processed to remove the skull and normalize image intensity across the field of view. An automated algorithm parceled the cortex into

discrete ROIs, based on local cortical curvature and neuroanatomical priors; additionally, it segmented major subcortical structures, based on gray/white matter contrast and neuroanatomical priors. FreeSurfer output included volumetric representations of all cortical parcels and subcortical segmented structures, as well as multiple representations of the cortical surface. Output from FreeSurfer was also converted for subsequent processing in AFNI and SUMA.

Because FreeSurfer output was in a different image orientation and voxel resolution than DSI data, it was necessary to align the two image spaces. For each dataset, the unweighted EPI volume (B0 image) was skull-stripped, then aligned to the skull-stripped, intensity-normalized T1 image output by FreeSurfer. Alignment was based on a 12-parameter affine transformation, computed using FSL's FLIRT. All transformed B0 images were visually inspected to check the alignment of ventricles, gyri, and sulci. The inverse of the previously computed affine transformation was subsequently used to align FreeSurfer-derived ROIs to the DSI data.

1.2.4 Diffusion Image Acquisition and Processing

Diffusion-weighted imaging data were collected twice for each of 7 participants with a 3-Tesla Siemens Tim Trio MR scanner and a 32-channel send/receive coil. We used a diffusion spectrum imaging (DSI) protocol with a twice-refocused spin-echo EPI sequence (TR=9916 ms, TE=156 ms, voxel size=2.40625 x 2.40625 x 2.4 mm, field of view=231 x 231 mm). This protocol scanned 257 diffusion directions over b-values ranging from 300 to 7000 s/mm². This sampling scheme represents anisotropic and isotropic diffusion in a four-dimensional space (three spatial dimensions plus time) referred to as “q-space” (Basser & Özarslan, 2009;

Callaghan et al., 1990). Images collected at higher b-values capture diffusion over longer time intervals than images at lower b-values, improving discrimination of crossing or “kissing” fibers. The cost of this improved diffusion contrast is decreased signal-to-noise ratio at high b-values (Tournier et al., 2004); however, empirical testing by our group and others has found that a mixture of high and low b-values produces tractography results which best reproduce known neuroanatomical features (Phillips et al., 2012; Wedeen et al., 2005).

No corrections for participant head motion, EPI distortions, or eddy currents were applied to the raw diffusion images. Although such corrections are commonly applied to diffusion-weighted MR images collected at lower b-values, no method currently exists for correcting data collected at high b-values, including DSI and multi-shell HARDI (Phillips et al., 2012). This is because current algorithms rely on global information about neuroanatomical structure across the volumes of an EPI dataset; however, those features are absent or severely degraded at high levels of diffusion weighting. An additional obstacle to correcting for head motion and image distortions is the spatial error introduced by image interpolation: even simple corrections, such as rigid-body alignments for motion correction, involve resampling and spatial blurring of the input data. The effects of such interpolation may go unnoticed in DTI-based tractography because of the limited angular resolution of the reconstructed data. In contrast, Gaussian spatial interpolation of DSI volumes may introduce spurious diffusion information into neighboring voxels, reducing the effective angular resolution in tractography results. We note that DSI protocols have been shown capable of producing highly reliable tractography results between sessions (Cammoun et al., 2012), despite the valid concerns about uncorrected image distortion and motion artifacts.

Generalized q-sampling imaging (Yeh et al., 2010) was used to reconstruct the raw diffusion data. Like q-ball imaging (QBI; Tuch, 2004) and the reconstruction scheme originally designed for DSI data (Wedeen et al., 2005), GQI belongs to the class of “model-free” reconstruction techniques, which employ Fourier transforms to estimate diffusion patterns from raw diffusion-weighted MR images. However, GQI is characterized by several differences from related model-free methods. The most significant difference is that ODFs in GQI are based on a spin density function (SDF) for each voxel, whereas QBI and DSI yield ODFs which are based on a probability density function (PDF). The PDF represents only proportions of the total spins diffusing in each direction within a given voxel and gives no information about the absolute number of spins represented. In contrast, the SDF calculated in GQI has been mathematically demonstrated to represent the absolute number of spins diffusing within a voxel (Yeh et al., 2010). GQI thus produces ODFs which have a meaningful physical relationship with those in other voxels. A second characteristic of GQI is that anisotropy is not usually expressed in terms of fractional anisotropy (FA; Basser & Pierpaoli, 1996; Jones, 2009), the metric prevalent in many other DWI methods. Instead, GQI employs a metric called quantitative anisotropy. While FA and generalized fractional anisotropy are directionally non-specific descriptions of the overall anisotropy (i.e., “pointedness” or “constrained-ness”) of diffusion in a voxel, QA is a directionally specific metric. QA can thus be estimated for multiple different diffusion orientations within a given voxel. Furthermore, in simulation experiments, QA has been shown to correlate with the physical volume of diffusion (Yeh et al., 2010), although this relationship has not yet been confirmed in diffusion phantoms or biological tissue.

Orientation distribution functions (ODFs) were reconstructed using DSI Studio (<http://dsi-studio.labsolver.org>; November 2011 release). The raw diffusion data (comprising 257 discrete

gradient directions) were resampled to represent 642 diffusion orientations, evenly distributed on a unit sphere. Possible fiber orientations were identified in each voxel by finding local maxima on the unit sphere representation. A maximum of 5 distinct diffusion orientations per voxel was allowed; this number was chosen to allow for the possibility of complex fiber crossings, although to our knowledge no evidence exists for crossings of more than 3 fiber tracts at a single point in mammalian brains. In practice, the reconstruction algorithm did not identify more than three possible fiber directions per voxel in any of the datasets analyzed. We employed a diffusion sampling ratio of 1.25.

1.2.5 Anatomical Region-of-Interest Creation

Anatomical ROIs for fiber-tracking and subsequent analysis were created from the FreeSurfer cortical parcels and subcortical segmentation described above. A total of 150 cortical ROIs were created for each brain (75 per hemisphere). Cortical ROIs were created from surface-format data, by dilating surface masks along the vector normal to the cortical surface at each node, to a depth of 1 times the cortical thickness below the gray/white matter boundary. This dilation is necessary because diffusion anisotropy falls off quickly in gray matter, and anisotropy estimates are less stable than in gray matter (Farrell et al., 2007). A small dilation of tractography seed or ROI masks thus serves to dramatically increase detection power for diffusion streamlines. Empirical testing indicates that our procedure dilates cortical ROIs by 1-3 voxels (a growth of 2.4-7.2 mm in the datasets presented here). Automatically-segmented subcortical ROIs were extracted from the FreeSurfer `aparc.a2009s+aseg` image.

Both subcortical and cortical ROIs were aligned to DSI data by applying the affine transformation previously computed for each dataset. This transformation involved a necessary loss of spatial resolution, due to the larger voxel sizes used in the DSI protocol. Nearest-neighbor interpolation was used for transforming ROI masks, in order to avoid the extensive blurring associated with trilinear interpolation. Subsequently, holes in transformed ROI masks were filled using a slice-by-slice procedure, in which any zero-valued voxels which were surrounded by at least 6 non-zero neighbors were also made non-zero.

1.2.6 Deterministic Tractography Methods

Fiber-tracking was performed with the deterministic tractography algorithm implemented in DSI Studio (Wedeer et al., 2008; Yeh et al., 2010). In general, deterministic tractography methods generate virtual fiber streamlines which represent possible fiber trajectories. Deterministic algorithms follow set decision rules at each step of a fiber's propagation; the resulting streamlines should be viewed as inferences of a fiber's presence, and not as veridical fibers. Furthermore, the streamlines generated by deterministic algorithms are not accompanied by estimates of their probability. In this respect, deterministic tractography differs from probabilistic algorithms, which estimate the probability of diffusion into a voxel as a combination of the diffusion signal strength and the accumulation of uncertainty created by tractography parameters. For example, using a large step size of 2 mm during fiber propagation creates more uncertainty than a step size of 0.5 mm, since larger step sizes decrease the algorithm's sensitivity to sudden changes in the trajectory of fiber bundles. While the accompanying confidence estimates represent an advantage of probabilistic tractography, this

method is less useful for making inferences of point-to-point connectivity in the brain: probabilistic methods typically produce volumetric “heat maps”, indicating the likelihood that fibers connecting two ROIs pass through each voxel. In contrast, the reliability analyses in the current study required estimates of connection density between ROIs; thus, deterministic tractography was employed.

As discussed in the introduction, the tractography algorithm used in DSI Studio requires the user to set several parameters that influence its operation. Analogous parameter decisions exist in other tractography software; when these parameter choices are not made apparent to the user, they may have been internally set in a software package. The most important of these are fiber or seed count, fiber length, tracking threshold, angle threshold, smoothing, and step size; the effects of variations in these parameters are investigated in subsequent analyses. Explanations of each of these tractography parameters can be found in Appendix A.

Yeh et al. (2010) outline the steps of the tractography algorithm used here. Briefly, seed points for fiber propagation were chosen randomly within the user-supplied seed mask. These seed points were chosen at sub-voxel resolution in continuous mm-based coordinates. The initial orientation of fiber propagation was randomly chosen from the peaks in the seed voxel. QA at the initial seed point was determined using trilinear interpolation from the 8 surrounding voxels in the plane of initial orientation. The algorithm then propagated the fiber in discrete steps (according to a user-set step size), choosing the orientation with the smallest angle of deviation from the current trajectory. This decision was mediated by the use of a momentum parameter, which represents the relative weights given to the incoming fiber trajectory and the trajectory at the next step: thus, a momentum value of 0.8 would yield a new fiber trajectory which was the sum of unit vectors representing the current orientation and the next chosen orientation, weighted

respectively by 0.8 and 0.2. The effect of this momentum parameter is to smooth fiber trajectories. Fiber propagation terminated if the virtual fiber encountered 1) a situation in which any further steps were at an angle of deviation greater than the user-set angle threshold; or 2) the QA values of any further steps were below the user-set tracking threshold.

1.2.7 Intercortical Connectivity Reliability Analysis

We sought to replicate Cammoun et al.'s (2012) finding of high between-session reliability in intercortical connection matrices. This analysis relies on FreeSurfer's anatomically-defined ROIs and examines connection density between every unique pair of ROIs (including interhemispheric connections). Although Cammoun et al. performed this analysis at multiple spatial scales, we employed a single grain size of 75 regions per hemisphere (150 total); this resolution is most similar to the 66/83-region resolution reported by Cammoun and colleagues (where 66 and 83 indicate the number of regions per hemisphere, before and after including subcortical ROIs).

Between-session reliability was assessed for each participant using multiple methods. First, we calculated the Pearson correlation coefficient between the connection matrices for the two sessions (Figure 2). Second, we assessed overlap in binarized connection matrices: for each session, we calculated the ratio of non-zero connections shared with the other session to the total of shared and unshared connections (Figure 3). The ratios for the two sessions were then averaged to produce the mean likelihood that a connection detected in one session would also be detected in a repeat session.

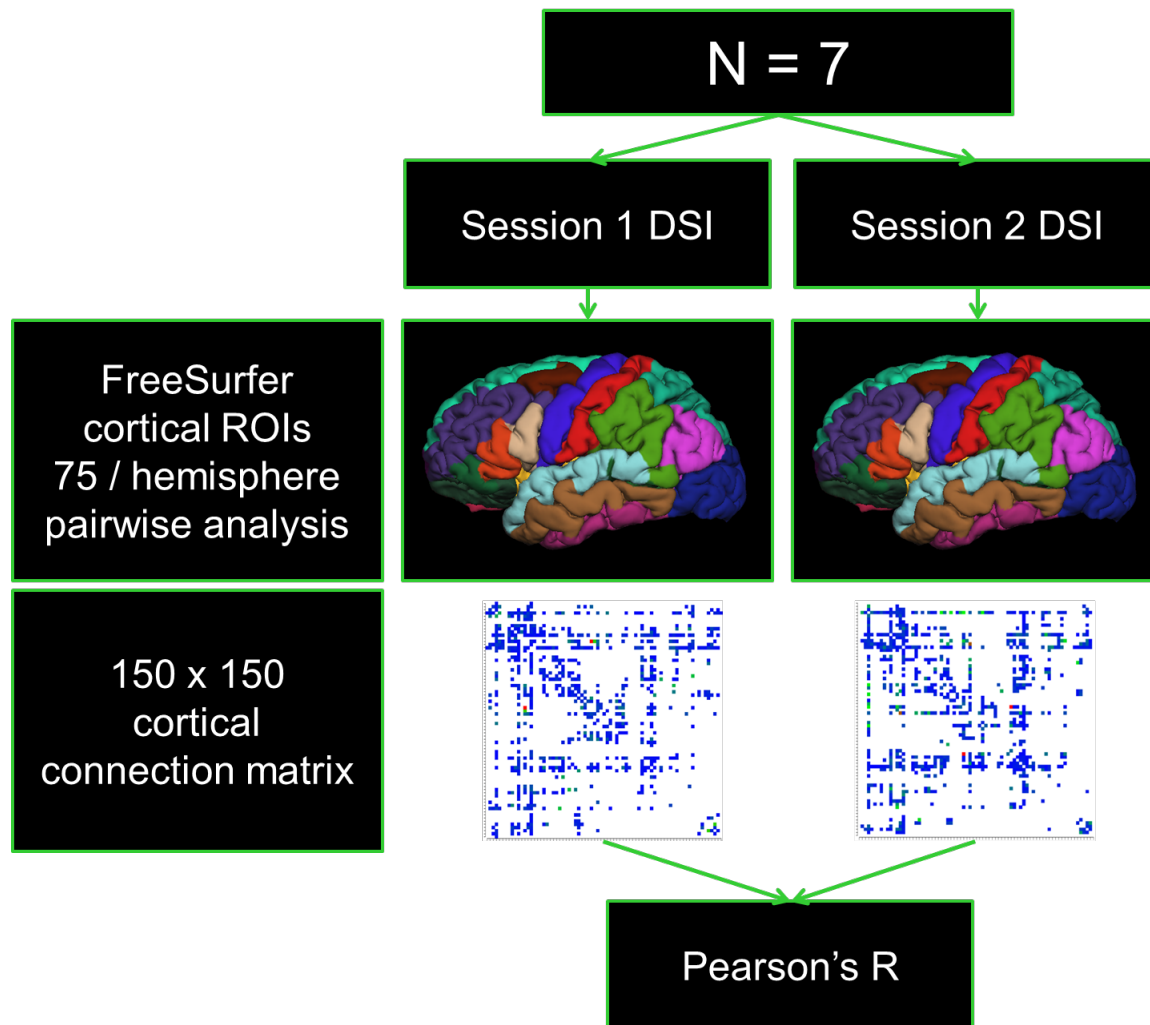
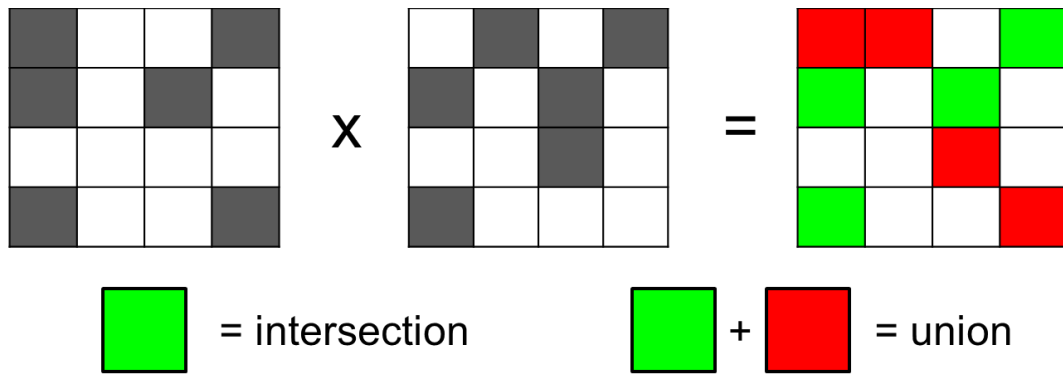


Figure 2. Intercortical connectivity analysis: Pearson's correlation metric.



Chance that a connection present in one session is present in the next:

$$\text{average} \left(\frac{\text{intersection}}{\text{session 1 all}}, \frac{\text{intersection}}{\text{session 2 all}} \right)$$

Figure 3. Intercortical connectivity analysis: overlap metric. Intercortical connectivity matrices are binarized (connection present/absent), and the proportion of repeatedly detected connections relative to total connections detected in a single session is calculated.

Tractography seed masks were created by aligning all FreeSurfer cortical ROIs to DSI data, then joining these ROIs into a single cortical mask. Tracked fibers were required to have at least one of two endpoints within this cortical mask. We used a fixed fiber count of 500,000 for all datasets, in order to facilitate comparison of connection densities across datasets. We used a maximum turning angle of 65 degrees, a smoothing value of 0.8, and a step size of 1 mm. These values were selected on the basis of prior studies from our group (Jarbo et al., 2011; Verstynen et al., 2011).

Additionally, a minimum length of 30 mm and maximum length of 200 mm were set. The lower length limit was chosen due to the objectives and granularity of the planned analysis: we aimed to detect only inter-regional connections traveling in white matter between FreeSurfer anatomical ROIs, and were not interested in self-connections or lateral connections within cortex. FreeSurfer ROIs range in surface area from approximately 1 cm² to more than 20 cm², and so short-range connections between adjacent cortical areas were expected to exceed the minimum distance. The upper length limit was chosen to set a generous margin for the detection of long-distance projection fibers.

The QA0 map, representing the magnitude of the largest QA value in each voxel, was used as the basis for anisotropy thresholding. The tracking threshold was initially set for one participant on the basis of values used for the same dataset in published analyses by our group (Jarbo et al., 2011; Verstynen et al., 2011). We then created a histogram of QA0 values from that dataset and found the z-score corresponding to the chosen threshold. The tracking threshold for the participant's second dataset and for all other datasets were determined by finding the QA value corresponding to the same z-score.

1.2.8 Targeted Tract Reliability Analysis

The analysis described in the preceding section (“Intercortical Connectivity Reliability Analysis”) provides an efficient method for assessing global brain connectivity with nearly whole-brain seeding. Moreover, it serves as a replication test of Cammoun et al.’s (2012) recent report on the reliability of DSI-based tractography. However, we sought to go beyond these findings, with three main objectives. First, we hoped to determine whether targeted tracking of specific white matter structures could improve between-session reliability, relative to the global tractography employed in the intercortical connectivity analysis. Second, we hypothesized that targeted tracking would increase the detection power of tractography, generating more streamlines than global tracking for a given fiber tract. Finally, we wished to investigate the repeatability of fiber-tracking results at the resolution of a single voxel, both in terms of tract termination fields and tract bodies.

Three prominent white matter structures were chosen for targeted tracking and analysis: the corticospinal tract, the arcuate fasciculus (anterior portion of the superior occipito-frontal fasciculus), and the interhemispheric projections passing through the corpus callosum. The corticospinal tract projects between spinal ganglia and the neocortex, particularly primary motor and pre-motor areas (Dum & Strick, 2002; Nieuwenhuys et al., p. 849). The arcuate fasciculus (AF) is a major long-distance projection pathway linking inferior temporal areas with prefrontal areas (Nieuwenhuys et al., p. 627-8). This tract, also referred to as the anterior branch of the superior longitudinal fasciculus, is implicated in speech and other forms of language processing, although active debate continues about its precise functional role (Bernal & Ardila, 2009; Catani & Mesulam, 2008). Finally, long-range projection fibers passing through the corpus callosum

form the principal path of communication between homotopic and heterotopic brain areas of the two hemispheres (Jarbo et al., 2011). Given the diverse topology of callosal projections, the importance of this fiber pathway is not limited to one or a few functional domains. The severing of callosal projections in split-brain patients has been associated with an array of curious behavioral and cognitive phenomena (Bloom & Hynd, 2005; Gazzaniga, 2000).

We targeted these structures for both theoretical and practical motivations: all three are of general interest to cognitive neuroscience researchers, have been investigated using such “gold-standard” techniques as microdissection and tracer labeling in primates (Dum & Strick, 2002; Killackey et al., 1983; Martino et al., 2010), could be located easily across individuals and datasets, and have been previously investigated by our research group (Fernandez-Miranda et al., in review; Jarbo et al., 2012; Verstynen et al., 2011). Additionally, the targeted tracts display morphological diversity, with both fanning and compact segments, and with principal fiber directions along each of the three cardinal axes. We felt that such morphological diversity was important for the generalizability of the findings presented here, to discount the possibility that our DSI protocol and hardware array were more sensitive to some tracts than to others.

Regions of interest were created for the three targeted tracts in each participant’s first DSI session; these ROIs were then spatially aligned to the second DSI session. Ipsilateral fibers of the left corticospinal tract were tracked using the method of Verstynen et al. (2011). Tracking was seeded in a cortical mask; virtual fibers were required to have one endpoint in this mask. Fibers were also required to pass through an ROI in the left cerebral peduncle and finally through an ROI located in the spinal cord, below the divergence of cerebellar fibers. Left AF fibers were seeded in a mask of temporal cortical areas, including inferior, middle, and superior temporal gyri; inferior, superior, and transverse temporal sulci; the temporal pole; lateral occipito-

temporal gyrus/sulcus; fusiform gyrus; and parahippocampal gyrus. Fibers were required to pass through a large rectangular prism ROI at the temporo-parietal junction, centered around the prominent, curving white matter pathway at this point; this ROI was manually placed for each dataset. Finally, fibers were required to terminate in a mask of pre-frontal cortical areas, comprising the frontomarginal gyrus/sulcus, inferior frontal gyrus (opercular and triangular parts), inferior frontal sulcus, middle frontal gyrus, superior frontal sulcus, lateral orbital sulcus, and precentral sulcus. Callosal fibers were seeded in gray matter of the left hemisphere, required to pass through a hand-drawn mask of the corpus callosum, and required to end in the contralateral gray matter. Seed and target masks for callosal tracking included not only cortical gray matter, but also the thalamus, amygdala, and basal ganglia.

An angle threshold of 70 degrees was used for targeted tracking, as Verstynen et al. (2011) previously used these values to investigate somatotopy in the corticospinal tract, and because the arcuate fasciculus and callosal projections were expected to exhibit similar amounts of curvature. Anisotropy threshold, smoothing, fiber length criteria, and step size were set as in the intercortical connectivity analysis. We generated 100,000 fibers for each tract of interest. To compare the reliability of targeted vs. global tracking, we also performed whole-brain tractography for all datasets without any ROI constraints. For whole-brain tracking, 500,000 fibers were generated per dataset; all other parameters were set as in the intercortical connectivity analysis.

1.2.9 Targeted vs. Global Tracking: Region-to-Region Analysis

We first analyzed targeted tracking results using the ROI-based method of the intercortical connectivity analysis. For each pair of tracts (session 1 and session 2), we computed overlap, correlation, and detection power metrics. Overlap was defined as between-session overlap of binarized connection matrices (i.e., the average likelihood of repeating a non-zero connection within subject, between sessions). The correlation metric was a pairwise Pearson's correlation of region-to-region fiber density values between each participant's two DSI sessions. As above, correlations of fiber density were computed using only the density values for connections which were non-zero in both of a participant's DSI sessions. Detection power was assessed by the number of unique connections detected in common across two sessions. These three metrics thus communicated complementary information: the likelihood of finding a connection to be reliably non-zero, the reliability of a connection's magnitude if repeatedly detected, and the absolute number of connections repeatedly detected.

Next we compared region-to-region connectivity results for targeted vs. global tracking approaches. We extracted fibers of interest from the whole-brain tracking files using the same ROI constraints employed in targeted tracking. For each participant, all tracts (3 structures x 2 conditions: global vs. targeted) were spatially aligned with the first session's DSI data using a previously-computed affine transformation. We then computed overlap, correlation, and detection power metrics for global tracking results. The effect of tracking condition on overlap and detection power was assessed using analyses of variance (ANOVAs) with fixed effects of tracking condition (global vs. targeted) and fiber tract (AF, callosal fibers, and corticospinal tract), as well as a random effect of participant. We predicted that targeted tracking would be

associated with greater overlap in binarized connection matrices than global tracking results; we also predicted that targeted tracking would be associated with greater detection power. To directly compare the between-session correlations of connection density values based on targeted vs. global tractography, we used Fisher's test for the difference of two correlation coefficients. This test was repeated for each participant, and the pattern of group results was qualitatively assessed. We predicted that between-session correlations would be higher under targeted tracking conditions than global conditions.

1.2.10 Targeted vs. Global Tracking: Voxelwise Analysis

Next, we investigated tractography reliability at the scale of a single voxel. The input data for these analyses were full-tract and endpoint-only density maps, in which the value of each voxel indicated the number of virtual fibers passing through or terminating within it. Full-tract density maps were used to compare the reproducibility of a tract's body between sessions, while endpoint density maps were used to assess the reproducibility of fiber termination fields.

As before, we rated each tract for each participant on overlap, correlation, and detection power; however, these metrics were slightly different for the voxel-wise analysis than for the region-to-region analyses described above. First, correlations were based on fiber counts in voxels of uniform size, count and density metrics were functionally equivalent. Second, it is important to note that by basing correlations on density maps, we did not strictly assess point-to-point connectivity. Density maps simply indicate how many of the fibers that connect seed and target ROIs occupy a given voxel; they do not indicate the specific seed/target voxels connected. This shorthand measure of connectivity was more computationally feasible than computing fiber

counts between every voxel and every other voxel in the brain; moreover, it allowed us to investigate the effects of spatial smoothing to address partial volume effects (see below). Between-session overlap was computed by calculating for each session the ratio of non-zero voxels in the between-session set intersection to the total count of non-zero voxels in that session, then averaging the two ratios. Thus, this overlap measure indicated the average likelihood that a voxel which was non-zero in one session's endpoint density map was non-zero in the second session. Correlation tests were performed for over the intersection of non-zero voxels in the two sessions' density maps. Finally, we assessed detection power in terms of the number of non-zero voxels shared between sessions in density maps. As above, we compared these metrics for targeted vs. global tracking results.

We anticipated that partial volume effects would have a sizable and adverse impact on the results of voxel-wise tests, as misalignments of just a millimeter or two could throw off endpoint density measures considerably. We addressed this problem by additionally investigating the effects of Gaussian spatial smoothing on correlation, overlap, and detection power metrics. We applied smoothing kernels of 2, 4, 6, 8, 12, and 20 mm full-width half-maximum (FWHM) to full-tract and endpoint-only density maps. We note that smoothing values of 4-6 mm are the most relevant for the reliability analyses reported below, as this range would address spatial misalignment of approximately 1 voxel in any given direction. Residual spatial error after affine-based alignment was not expected to exceed 1-2 voxels.

We compared targeted vs. global tracking results using the same statistical methods described in the region-to-region analysis. The effect of tracking condition on overlap and detection power metrics was assessed using ANOVAs with fixed effects of tracking condition and fiber tract, as well as a random effect of participant. As in region-to-region analysis, we

predicted that targeted tracking would be associated with greater between-session overlap and greater detection power. Conditional differences in the reliability of fiber density estimates were again assessed by testing the difference between correlation coefficients under targeted vs. global tracking conditions. We again predicted that between-session reliability of fiber density estimates would be more reliable under targeted tracking conditions.

1.2.11 Identifying Sources of Between-Session Error in Tractography Results

We next performed a series of follow-up analyses to identify potential sources of error that could adversely affect between-session reliability. Specifically, we investigated 1) differences resulting from random seed generation during tracking; 2) variability related to differences in tractography parameter settings between sessions; 3) error related to spatial misalignment of observations; and 4) acquisition-related error, encompassing factors such as participant head motion and image distortions.

The first of these factors, which we call reseeding error, does not involve between-session variability per se; rather, it refers to the possibility that differences in random seed points between runs of the tractography algorithm could lead to the generation of different fiber sets, ultimately causing differences in interregional connection estimates. This problem may appear even greater if we consider that the tracking software employed (DSI Studio; <http://dsi-studio.labsolver.org>) generates seed points at sub-voxel resolution, implying a practically infinite number of possible starting points for fiber propagation. However, many of the fiber streamlines generated through extensive sub-voxel seeding are redundant with one another, following parallel trajectories with minimal spatial separation. Thus, at the resolution of our analyses, these

fibers would be practically identical, and small differences between them would be irrelevant to between-session analysis. In theory, a sufficiently large fiber or seed count should saturate the space of non-redundant fiber trajectories. Given the large fiber counts which we employed, we predicted that reseeding error would not be shown to adversely affect between-session reliability. To test this hypothesis, we performed reseeding of tractography at multiple fiber counts; detailed procedures are described below.

The second error source which we addressed was the sensitivity of tractography results to variations in tracking parameters. As mentioned above, key parameters in deterministic tractography include anisotropy threshold (hereafter referred to as tracking threshold), maximum turning angle, minimum and maximum fiber lengths, step size for fiber propagation, smoothing, and the number of seeds or fibers generated by the tractography algorithm in a single run. These parameters are defined in greater detail in Appendix A. Theoretically, sub-optimal parameter choices could degrade the between-session reliability of tractography results. For example, error in equilibrating tracking threshold could create variability in tractography results, or the use of an excessively small step size could make tractography more sensitive to noise sources that preferentially affect one of the two datasets compared. To our knowledge, no study to date has investigated the effect of systematic variations in tractography parameters on the results. We analyzed the density and morphology of 3 representative white matter structures (corticospinal tract, arcuate fasciculus, and callosal projection fibers) across a large combination of parameter values, attempting to identify the range of each variable which produced mutually consistent and anatomically correct results. The detailed procedure of this analysis is given below.

The third error source we investigated was spatial misalignment of datasets. We reasoned that spatial error could occur in multiple situations. One form of spatial error is variability

between sessions in the cortical parcellation and subcortical segmentation performed by FreeSurfer. FreeSurfer processing was performed independently for the two datasets collected for each participant, using the T1-weighted anatomical image collected in each scanning session. Although brain morphology was unlikely to have changed significantly for any participant between scans, differences in magnetic field homogeneity, hardware state, or other factors could have caused differential patterns of image distortion between scans, potentially changing the curvature and contrast information employed by the FreeSurfer recon-all pipeline. Moreover, Cammoun et al. (2012) report that FreeSurfer processing is not deterministic, yielding small differences when run multiple times on the same dataset.

Another form of spatial error is imperfect alignment of datasets for voxel-by-voxel comparisons. Our initial analysis of between-session reliability is not affected by this form of spatial error, as correlations were performed between the pairs of independently-derived FreeSurfer anatomical ROIs, as described above. However, spatial transformation error is highly relevant to the voxel-by-voxel comparison of endpoint fields for the specific tract analysis detailed above. Like the T1-weighted anatomical images, the $b=0$ EPI images used for alignment of each participant's two DSI datasets may suffer from different non-linear distortion patterns, which render perfect alignment with affine transformation methods impossible. The best overall transformation for a given pair of datasets can produce a misalignment of several millimeters for some specific structures, reducing voxel-by-voxel correlations of fiber density. It is difficult to directly quantify the effect of spatial error on tractography results; however, we attempt to assess the magnitude of spatial error in several indirect analyses, described below.

Finally, acquisition-related error could represent a major factor adversely affecting the reproducibility of tractography results. This category combines measurement error and

physiological noise, including non-linear EPI distortions, eddy currents, participant head motion, cardiovascular effects, and respiration. In the temporal sequence from acquisition to tractography, these causes are primary, thus potentially affecting all subsequent processing and analysis. As discussed above (“Deterministic Tractography Methods”), the high b-values employed in DSI degrade global anatomical structure in EPI volumes, making estimation and correction of motion parameters impossible; for the same reason, EPI distortions and eddy currents cannot be addressed. Finally, the specific effects of physiological processes upon the diffusion signal upon the current dataset are unclear. However, respiration and cardiovascular processes are known noise sources in MRI generally, as they create magnetic field inhomogeneities and can contribute to motion artifacts. Physiological noise in DTI data has been previously associated with biologically invalid diffusion estimates, in the form of negative eigenvalues for tensor axes (Jones, 2009). Although the current study was unable to dissociate these acquisition-related errors from one another, we sought to characterize their aggregate effect upon data quality, relative to other error sources mentioned above. Detailed procedures for investigating each error source are described below.

1.2.11.1 Assessment of Reseeding Error

We assessed reseeding error by performing repeated tracking (two iterations) at each of 6 fiber counts, for all participants. This procedure was performed for intercortical connections, as in our initial reliability analysis. We predicted that connectivity results for the pairs of tracts would converge quickly and become identical at high fiber counts. Tracking was performed at

fiber counts of 100, 500, 1000, 5000, 10000, and 100000. Similarity was assessed by correlating the FreeSurfer ROI connection matrices for the two sessions, as in our initial analysis.

1.2.11.2 Parameter Sensitivity Analysis

For each parameter investigated, we aimed to identify the range of values which produced morphologically consistent and neuroanatomically valid results. We did this by orthogonally varying several tracking parameters, to produce multiple copies of a target tract. We then tested each tracking result against a set of neuroanatomical priors and eliminated failing results from consideration. Next we divided the surviving results into quartiles, based on their net accuracy score. We computed a similarity matrix for the surviving results and found the mean mutual similarity of results within each quartile.

Tracking threshold, angle threshold, and smoothing were parametrically and orthogonally manipulated. Step size was held constant at 1 mm, since the inclusion of this variable in parameter tests dramatically increased computing time and made testing unfeasible. Parameter tests were conducted in a single participant's dataset for the left corticospinal tract only, due to the computational intensity of the tractography demands associated with this analysis.. Five values each were sampled for angle threshold and smoothing; a range of ten tracking thresholds was sampled. Thus, tracking for each of the three tracts was performed 250 times.

The parameter sensitivity test was performed in two phases. In the first phase, 25 million random seed points were generated in each tracking job. The number of seeds (as opposed to a count of virtual fibers) was chosen as the stopping criterion for a tracking job, because extreme parameter values—particularly very high tracking thresholds and very low angle thresholds—

could reduce or entirely prevent the detection of virtual fibers. Tracks produced during this first iteration were then evaluated against a set of anatomical heuristics. These tests were based on prior neuroanatomical findings from invasive techniques, such as microdissection and tracer labeling. For each of the three test tracts, we compiled sets of “hit” and “false alarm” regions. The former set comprised brain areas known to be contacted by the tract in question, while the latter set comprised areas to which the tract did not project. Detailed descriptions of these heuristics for each test tract are given below.

We subtracted the false alarm rate for each virtual fiber tract from its hit rate to yield a composite accuracy score. Only tracts with a net accuracy of at least 0.5 were submitted to the second phase of testing. In this phase, a uniform total of 100,000 virtual fibers was generated for each fiber tract. This uniform count was chosen to allow comparisons of fiber density ratios across the results produced by different combinations of tracking parameters.

Ten FreeSurfer ROIs were designated as target regions for the corticospinal tract, on the basis of known white matter projections to these areas. The target ROIs encompass primary somatosensory and motor cortices, as well as dorsal and ventral premotor areas, which form the bulk of the corticospinal tract. We also included the posterior portions of the superior and middle frontal gyri, as well as the posterior part of the superior frontal sulcus, on the basis of these areas’ putative homologies to macaque areas F2 and F7 (Geyer et al., 2000). Similarly, a set of false alarm ROIs was compiled from brain areas known not to send projections into the corticospinal tract. The complete list of target and false alarm ROIs is given in Table 1.

Table 1. Target and foil regions for parameter sensitivity test.

<i>Hit regions (ipsilateral only)</i>	<i>False alarm regions (bilateral)</i>
Central sulcus	Amygdala
Paracentral gyrus and sulcus	Calcarine sulcus
Precentral gyrus	Corpus callosum
Postcentral gyrus	Hippocampus
Postcentral sulcus	Inferior temporal gyrus
Anterior superior frontal gyrus	Lateral occipito-temporal/fusiform gyri
Superior frontal sulcus	Medial occipito-temporal/lingual gyri
	Medial occipito-temporal/parahippocampal gyri
	Superior temporal gyrus/planum temporale
	Temporal pole
	Precuneus
	Intraparietal sulcus

1.2.11.3 Assessment of Spatial Error

Spatial differences in FreeSurfer anatomical ROIs were assessed in two ways. First, we addressed the possibility that randomness in FreeSurfer recon-all processing could cause differences in cortical parcellation or subcortical segmentation between runs (Cammoun et al., 2012). For each of 3 participants, we performed two iterations of the recon-all pipeline on the same T1-weighted anatomical image; ideally, we expected the results to be completely identical. We aligned the resulting ROI masks to the DSI data from the same session by applying the previously computed affine transformation, then compared overlap in the masks produced by the two FreeSurfer runs. Overlap was defined as above, by computing the average likelihood (between the two sessions) that a voxel found in the ROI mask would be found in the set intersection with the corresponding ROI mask from the other session. These results were interpreted as a measure of how reliably FreeSurfer would identify a given anatomical structure, given the same input.

Next, we addressed the possibility that differences in non-linear distortion between the two T1 images could create variability in FreeSurfer's definition of cortical and subcortical ROIs. Anatomical ROIs derived from two different sessions were aligned to the same session's DSI data, using the previously-computed affine transformation and nearest neighbor interpolation. Again we compared the overlap between the two sessions. These results were taken as a measure of FreeSurfer's consistency in assigning the same set of voxels to a given anatomical ROI, given different T1-weighted images that could be differentially affected by non-linear image distortions. However, this interpretation was complicated by the need to first align the two sets of anatomical ROIs to a common space, as well as the interpolation which accompanied each of

these alignments. Thus, this analysis was confounded with spatial transformation error, such as results from partial volume effects.

Spatial transformation error of this sort could particularly affect the voxel-wise correlation of fiber density in the specific tract analysis reported above. We expected that moderate amounts of spatial smoothing in fiber density maps would be largely effective in compensating for this error; however, smoothing could not be expected to be effective in specific areas or individuals with large alignment error. To gauge the presence and relative severity of spatial transformation error across participants, we assessed the quality of spatial alignment between each participant's pair of B0 images. This was done by spatially aligning the session 2 image to its session 1 counterpart, using the affine transformation computed above. We then computed voxel-wise correlation coefficients between the two images. We predicted that maximal correlation between the two B0 images would be obtained with a moderate amount of spatial smoothing (e.g., 1-2 voxels).

1.2.11.4 Assessment of Acquisition-Related Error

We assessed acquisition-related measurement error in multiple ways. First, we compared the distribution of QA values in each pair of QA0 maps (which were the basis for anisotropy thresholding in tractography). We reasoned that if QA estimates were reliable across sessions, the two distributions should heavily overlap. Furthermore, the degree of differences between histograms of the two distributions should provide a coarse metric of the error between the two datasets. We first converted each distribution into z-scores, such that their means were aligned at zero, and subtracting one session's histogram from the other's. By taking the sum of absolute

values of the remaining counts across histogram bins, we obtained a rough estimate for each participant of the similarity of the two QA0 maps. Although this method is simplistic, it avoids the confound of spatial transformation error introduced by voxel-wise comparisons.

Next, we examined between-session differences in anisotropy estimates on a voxel-by-voxel level for each participant. We reasoned that the difference between QA maps should be inversely related to the reliability of QA estimates. As in the subtraction of QA0 distributions described above, values in the maps were converted into z-scores; the session 2 map was then aligned with the session 1 map. For each participant, we computed the Pearson correlation coefficient across all voxels. We then compared between-session correlation of B0 images to that of QA0 images. We reasoned that any decrement in correlation for QA0 images, relative to B0 images, would indicate variability in anisotropy values (and thus, diffusion contrast) which could not be attributed to residual spatial alignment error.

Additionally, we assessed the spatial distribution of QA0 variability by creating maps of “masked-out” voxels: that is, portions of the QA0 map which survived the anisotropy threshold used in one session, but not the second. We reasoned that this voxel-wise analysis could identify datasets marked by less reliable QA estimates, or whose tracking thresholds had not been well equilibrated, eliminating true signal in at least one of the two sessions. Furthermore, this analysis indicated brain volumes that were particularly susceptible to variability in anisotropy measurements across individuals. We predicted that unreliable voxels would be particularly prevalent near tissue boundaries in the brain, due to true differences in anisotropy between tissue types, as well as the exaggerated visibility of head motion artifacts in these areas. While the spatial specificity of this analysis was desirable, we note that it is partially compromised by the possibility of spatial transformation error, as discussed above: apparent variability in QA

estimates between sessions could be due at least partially to poor spatial alignment of the maps to be compared.

1.3 RESULTS

1.3.1 Intercortical Connectivity Reliability Analysis

1.3.1.1 Between-Session Correlation of Connection Density

We first assessed the between-session correlation of connection densities among all unique pairs in the set of 150 FreeSurfer cortical ROIs. Initially, we calculated Pearson's correlation coefficients for each participant over the set of connections which were non-zero in one or both sessions (Table 2). These results were markedly lower than the correlation coefficients found by Cammoun et al. (2012): at their analogous resolution of 83 ROIs per hemisphere, the authors report single-subject correlation coefficients with a mean of approximately 0.98, indicating almost perfect correlation between the connection densities in intercortical connection matrices obtained for the two sessions. In contrast, we found correlation coefficients which ranged from 0.14 to only 0.53, with a mean of approximately 0.38 across participants.

Table 2. Within-subject, between-session reliability of intercortical connectivity matrices.

Participant	R (between-session union)	R (between-session intersection)	Overlap of binarized connection matrices
1	0.41	0.95	0.84
2	0.40	0.92	0.77
3	0.14	0.93	0.65
4	0.37	0.93	0.77
5	0.39	0.94	0.70
6	0.45	0.95	0.77
7	0.40	0.94	0.77
Mean	0.38 (0.11)	0.94 (0.02)	0.75 (0.06)

We asked whether the difference in our results and those of Cammoun et al. (2012) could relate to the thresholding procedure mentioned by those authors. This procedure excluded cells in which ROI pairs were relatively sparsely connected (by fewer than two fiber bundles). Additionally, Cammoun et al. excluded a number of interhemispheric projections from frontal lobe regions for unstated reasons. We speculated that these steps had eliminated a great deal of noise in their correlations, possibly by excluding cells which had failed to replicate between sessions. Indeed, scatterplots of session 1 vs. session 2 connection densities in the current analysis revealed a large number of connections that were present in one session but failed to be replicated in the second dataset from the same participant. We refer to such connections hereafter as “one-off” connections, and we note that these cases probably represent a mixture of false positive and false negative results. The frequency of these one-off cases can be derived from the overlap metrics in Table 2, which indicate the average likelihood for each participant that a non-zero connection in one session would be non-zero in the other session. The average probability

of repetition for a single participant was 0.75; thus, the average frequency of one-off cases was about 0.25 of all non-zero connections reported for a single session. A paired t-test indicated no significant effect of session upon the number of one-off connections detected [$t(6)=-1.88$, $p<0.11$], discounting the possibility of systematic effects of scanning sequence upon data quality. A one-way ANOVA also revealed no effect of subject upon the number of one-offs, suggesting that the rate of failure of replication was relatively homogeneous across individuals [$F(6,7)=0.14$, $p<1$].

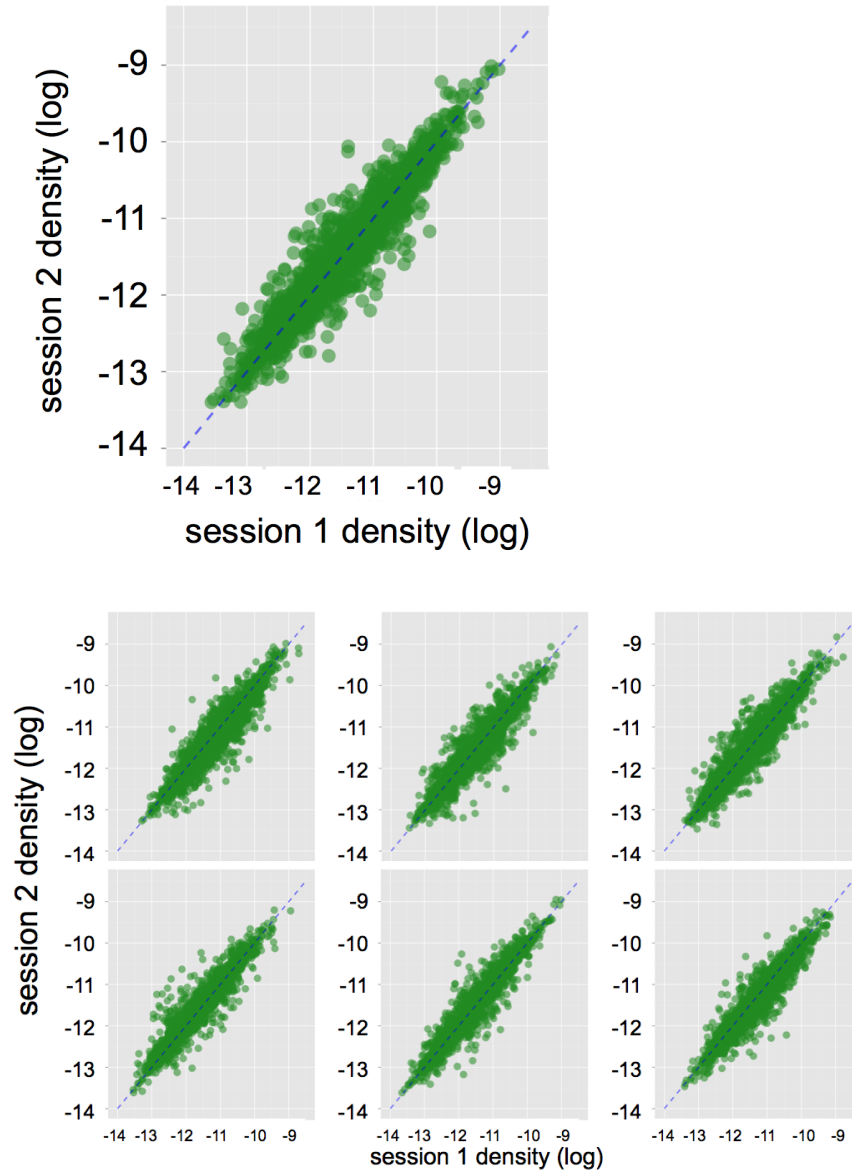


Figure 4. Between-session comparison of log-transformed intercortical connection densities. Top: Scatterplot of log-transformed intercortical connection density values for two DSI sessions from participant 1. Each point represents a unique connection between two anatomically-defined ROIs. The dotted blue line, oriented at 45° to the x- and y-axes, represents

In contrast, cells which were non-zero in both sessions appeared to correlate well. The scatterplots of Figure 4 show a highly consistent relationship between session 1 and session 2 connection density values for each participant when one-off connections were excluded from consideration. We confirmed this relationship by re-computing correlations over the intersection of non-zero connections in paired datasets (i.e., the set of connections in common between each participant's two sessions). As expected, focusing on connections which were repeatedly non-zero significantly improved the correlation of connection densities, yielding Pearson's coefficients ranging from 0.92 to 0.97 (Table 2). These results are comparable with the results of Cammoun and colleagues, indicating high reliability in the connection densities of overlapping cells of the two sessions' connection matrices.

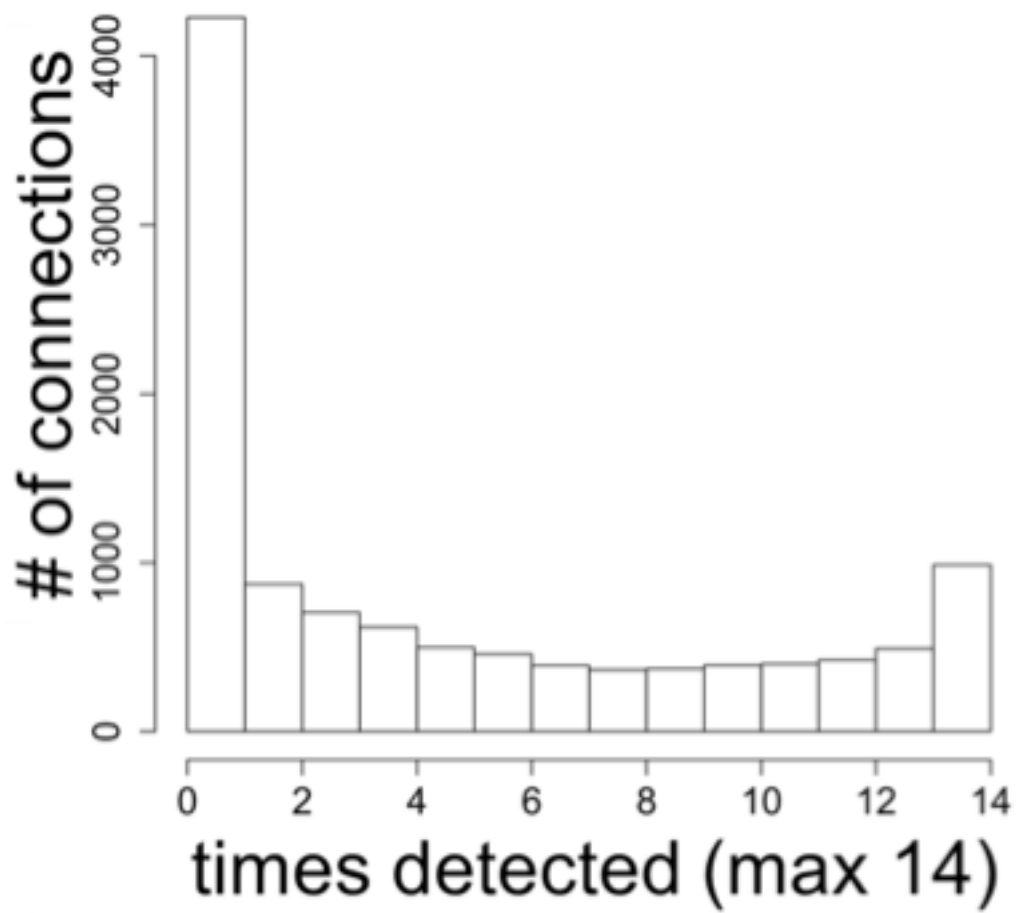


Figure 5. Detection frequency for intercortical connections. A total of 11175 possible pairwise connections exist in the ROI matrix. Of these, 987 were detected in all 14 DSI datasets.

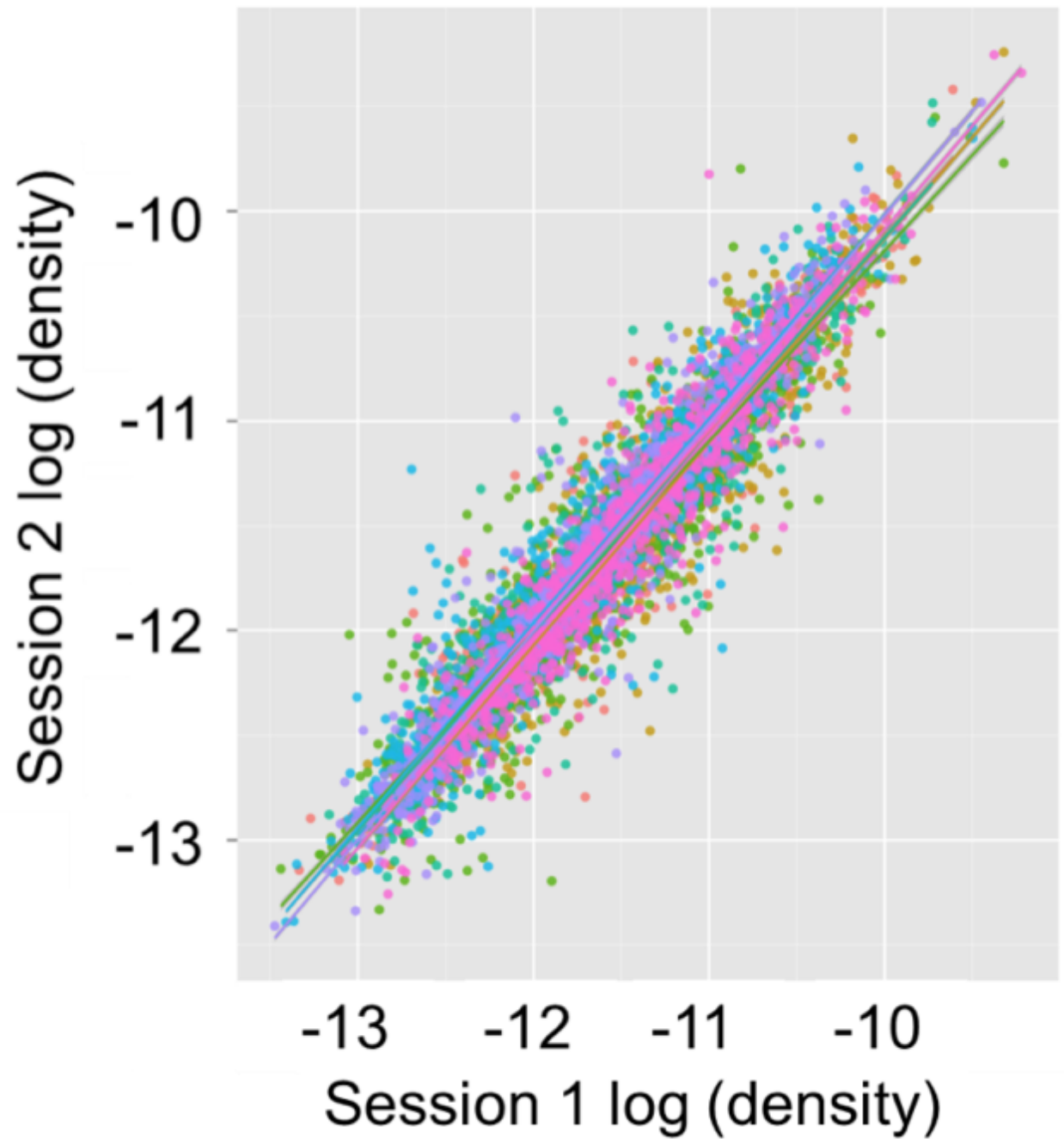


Figure 6. Scatterplot of density values for connections detected in all datasets. Each color represents a different participant. Values are log-transformed. Session 1 density values were a strong linear predictor of session 2 values.

1.3.1.2 Reliability Across Individuals and Sessions

Of course, the results described above merely imply consistency over time within a given individual; they do not indicate the degree of replicability across individuals. To address this question, we summed binarized connection matrices from all 14 datasets to yield the overall number of detections for each cell, both within and between subjects. Figure 5 shows a histogram of these detection counts. A total of 987 connections replicated across all 14 datasets; this number represented, on average, a proportion of 0.24 (std. dev. = 0.02) of the total non-zero connections detected in a given dataset. This figure should not be interpreted to suggest that only 24% of the intercortical connections reported in a given dataset are true connections. Rather, the remaining 76% of connections are likely to reflect a combination of misses in some datasets, false alarms in others, and real individual differences in anatomical connectivity. A scatterplot restricted to these most often detected connections indicated that their between-session reliability remained extremely high (Figure 6).

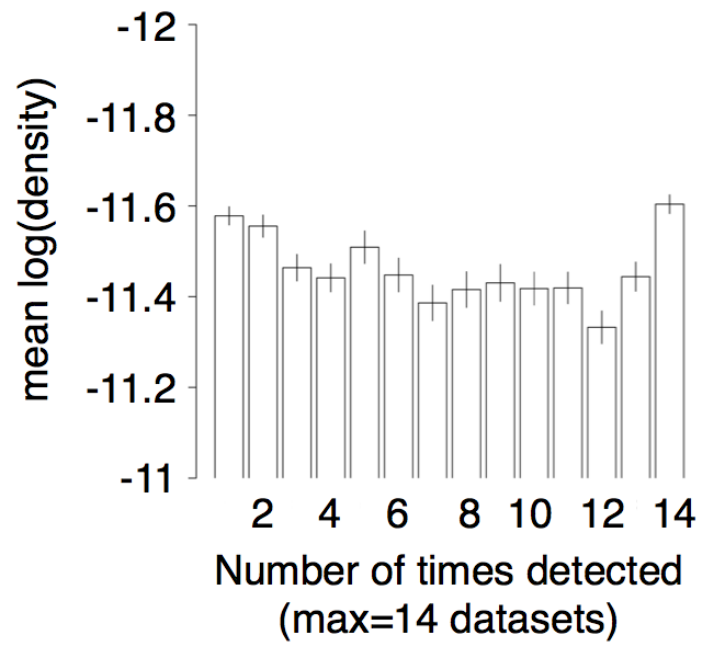


Figure 7. Relationship between connection density and frequency of detection.

We further asked how the probability of replicating a given connection related to its strength, as measured by the density metric of Cammoun et al. (2012). We predicted that stronger connections would be associated with a greater number of detections across datasets than weaker connections. In order to test this prediction, we first calculated the mean fiber density across datasets for each cell in the intercortical connection matrix. (We note that means were based only on the set of non-zero values for each connection; thus, the mean for a connection detected in all 14 datasets is based on more data than the mean for a connection detected in just one or two datasets.) The mean of these values for each level of detection (1 to 14 times) is shown in Figure 7. A linear regression model indicated a significant but very modest relationship between the number of detections and connection density [coefficient=0.005; $t(8328)=2.5$, $p<0.011$; adjusted $R^2=0.0007$], according to which denser connections had only slightly more detections than sparser connections. Although this result reached statistical significance, its small size (coupled with visual inspection of Figure 7) suggests that higher connection density is not a major predictor of the likelihood of replication within- and between-subjects. The difficulty of using connection density as a heuristic to separate reliable from unreliable connections is made especially clear by Figure 8, which plots distributions for each of three different connection types: those that were detected in all 14 datasets, those that were successfully replicated within-subjects (but not across all datasets), and one-off connections. We note that despite variability in the height of each distribution, the three connection types have similar means and widths in all 14 datasets.

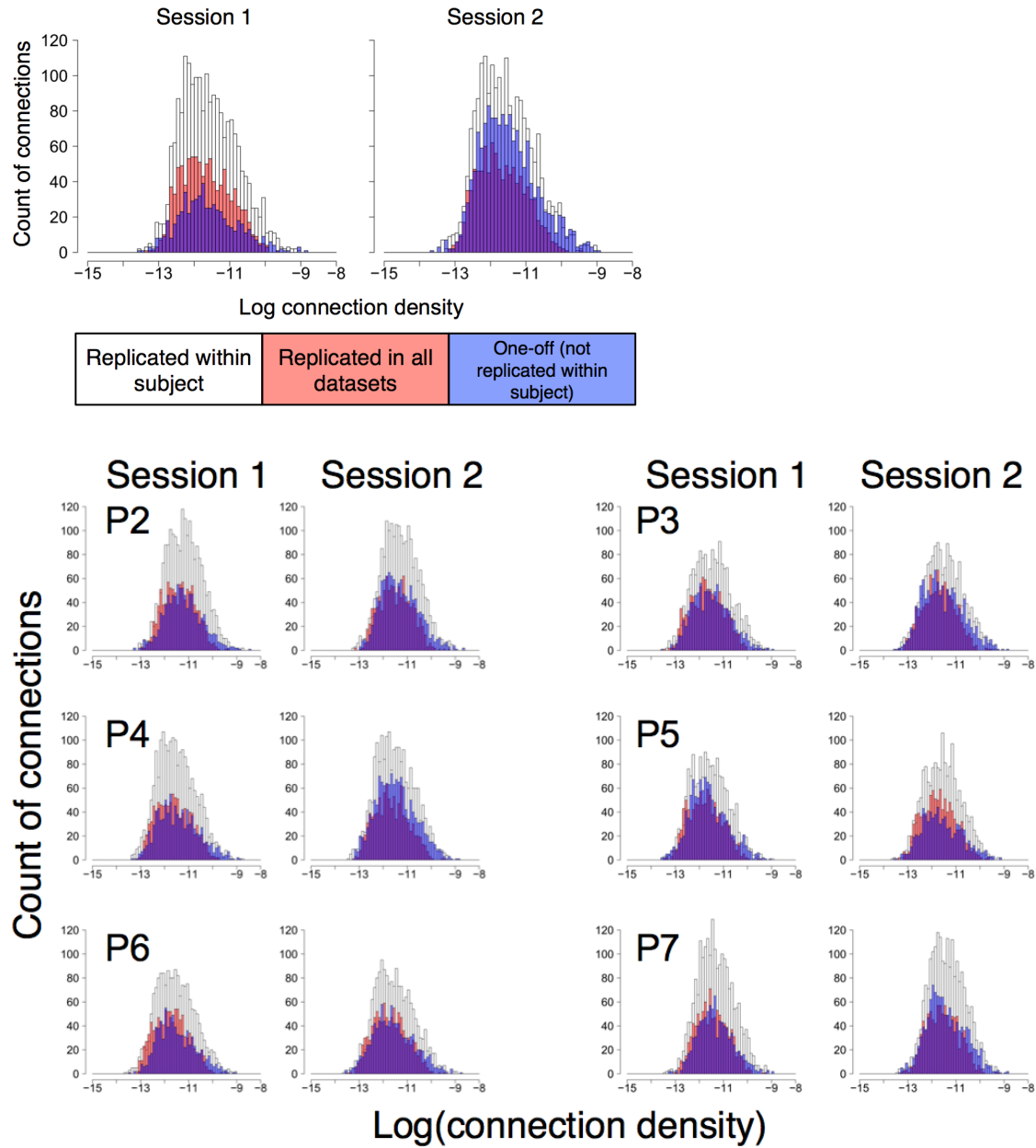


Figure 8. Distribution of connection densities by connection type. One-off connections are plotted in blue, connections replicated in all 14 datasets are plotted in red, and connections replicated within-subject (but not across all datasets) are plotted in white. Purple bars indicate overlap between red and blue distributions.

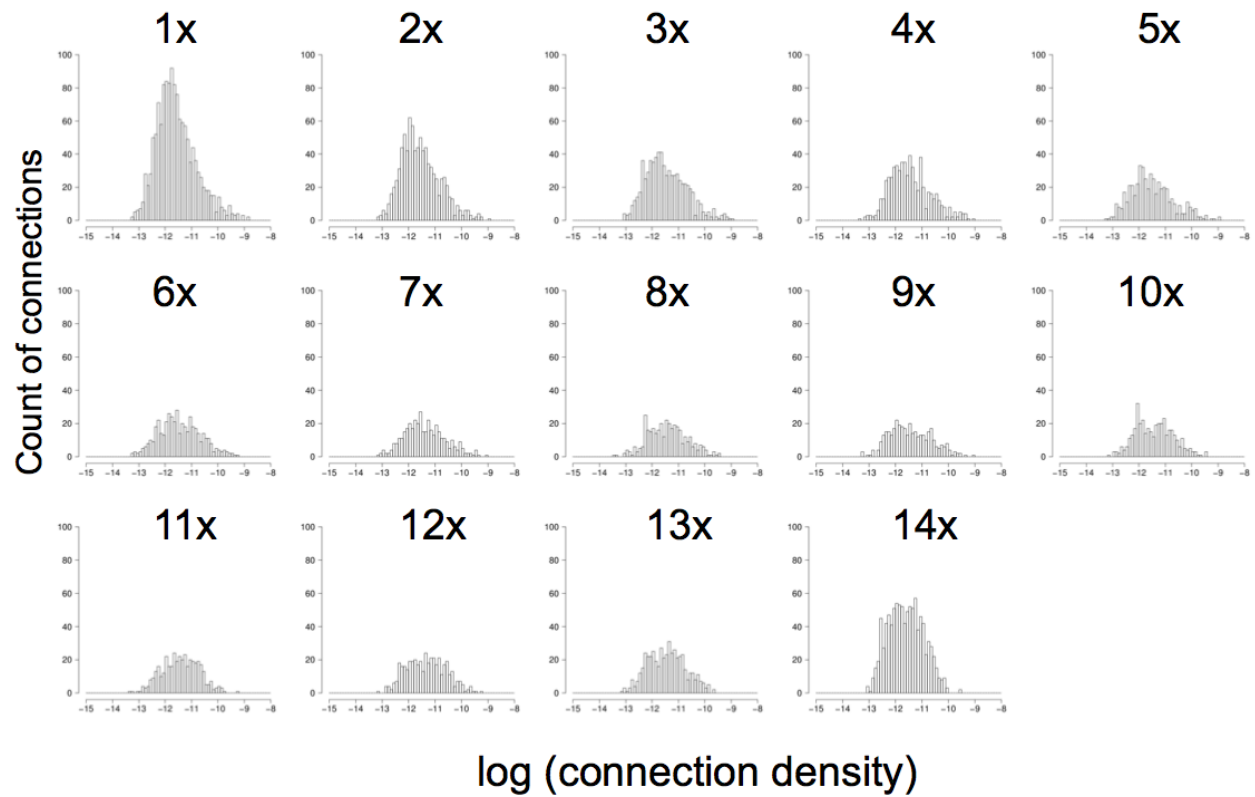


Figure 9. Histograms of connection density by frequency of detection. In all histograms, the x-axis represents the natural logarithm of fiber tracts' connection density, calculated according to the method of Cammoun et al. (2012). The y-axis represents the count of connections detected within each participant at each density level.

We next sought to characterize in more detail those connections detected in all 14 datasets, which we refer to hereafter as most-often-detected (MOD) connections. Of all fiber pathways detected in a given dataset, these MOD connections may be the most likely to represent veridical white matter tracts, due to their high consistency across time and between individuals. In contrast, we predicted that if random noise in DSI datasets created false fiber pathways, they would be inconsistently detected between- and within-subjects, due to variability in spatial location and trajectory. A total of 987 connections were detected in all 14 datasets; Figure 10 characterizes these connections in terms of their trajectory (intrahemispheric vs. interhemispheric) and their left-right symmetry. Intrahemispheric connections constitute the majority of these MOD connections; within the set of intrahemispheric connections, the majority of those found in one hemisphere are independently detected in the other hemisphere (298 connections with opposite-hemisphere homologues vs. 135 left-only and 182 right-only connections). Of 72 interhemispheric connections, a minority of 6 connections were homotopic, extending from one area to the homologous area in the opposite hemisphere. Thirty-eight connections were heterotopic and symmetrical: they extended from one area to a non-homologous area in the opposite hemisphere, and the left-right mirror image of that connection was also present. (Collapsing across hemisphere, these connections thus represent 19 unique fiber pathways.) A total of 30 interhemispheric connections were heterotopic and asymmetrical.

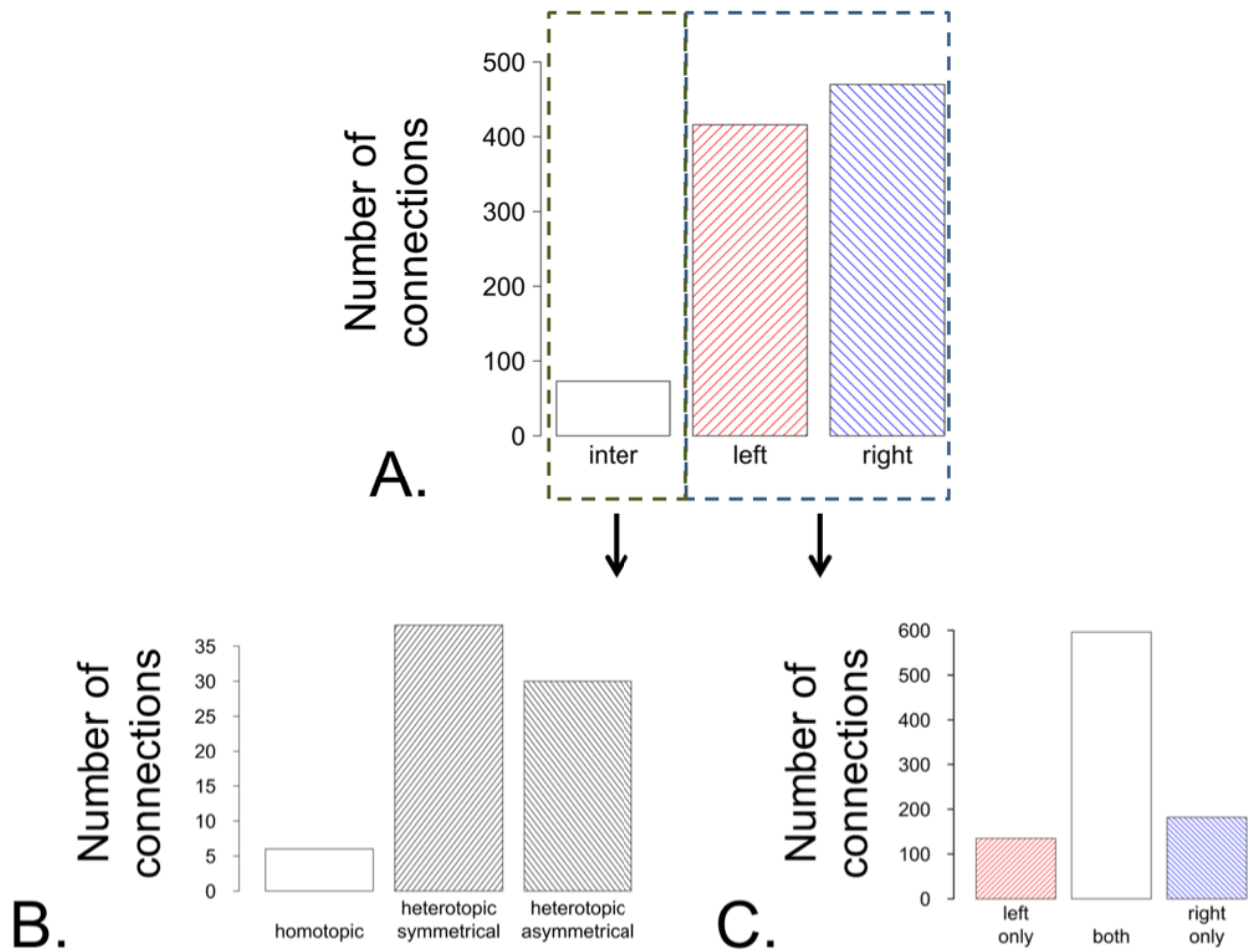


Figure 10. Type and symmetry of interhemispheric connections replicated in all datasets.

Plot A: counts of interhemispheric, left-hemisphere intrahemispheric, and right-hemisphere intrahemispheric connections. Plot B: counts of homotopic, symmetrical heterotopic, and asymmetrical heterotopic connections within the set of interhemispheric connections. There are 38 symmetrical connections; thus, 19 connections in a single hemisphere are mirrored in the opposite hemisphere. Plot C: counts of left-hemisphere-only, right-hemisphere-only, and symmetric (i.e., present in both hemispheres) connections composing the set of intrahemispheric connections (middle and right bars in Plot A). In a single hemisphere, 298 connections have a homologue in the opposite hemisphere.

We next investigated the spatial distribution and network structure of MOD connections. We first identified the cortical areas with the highest degree (Rubinov & Sporns, 2009), a graph theory metric which here indicates the sum of non-zero MOD connections projecting to each cortical area. Degree values ranged from a minimum of 1 to a maximum of 67 connections. The mean degree was 13.8 (std. dev. of 12.0), with a median degree of 11. Results were highly similar across hemispheres: of the top 10 left-hemisphere areas of highest degree, 7 of their contralateral homologues were in the corresponding right-hemisphere list (Table 3). The cortical wall had the highest degree of connectivity in each hemisphere, possibly due to the inclusion of fibers terminating in the thalamus. Additional areas of high degree included bilateral middle and superior frontal gyri, superior parietal lobules, orbital gyri, and superior temporal sulci. Anterior cingulate cortex, the middle temporal gyrus, and the lateral superior temporal gyrus were also highly connected (Table 3, Figure 11). The node-to-node topology of the MOD connection network is illustrated in Figures 12–14, plotted separately for inter- and intra-hemispheric connections. These figures provide supplementary information to Table 3 and Figure 11, including the presence of homotopic intercortical connections among several of the high-degree regions listed in Table 3.

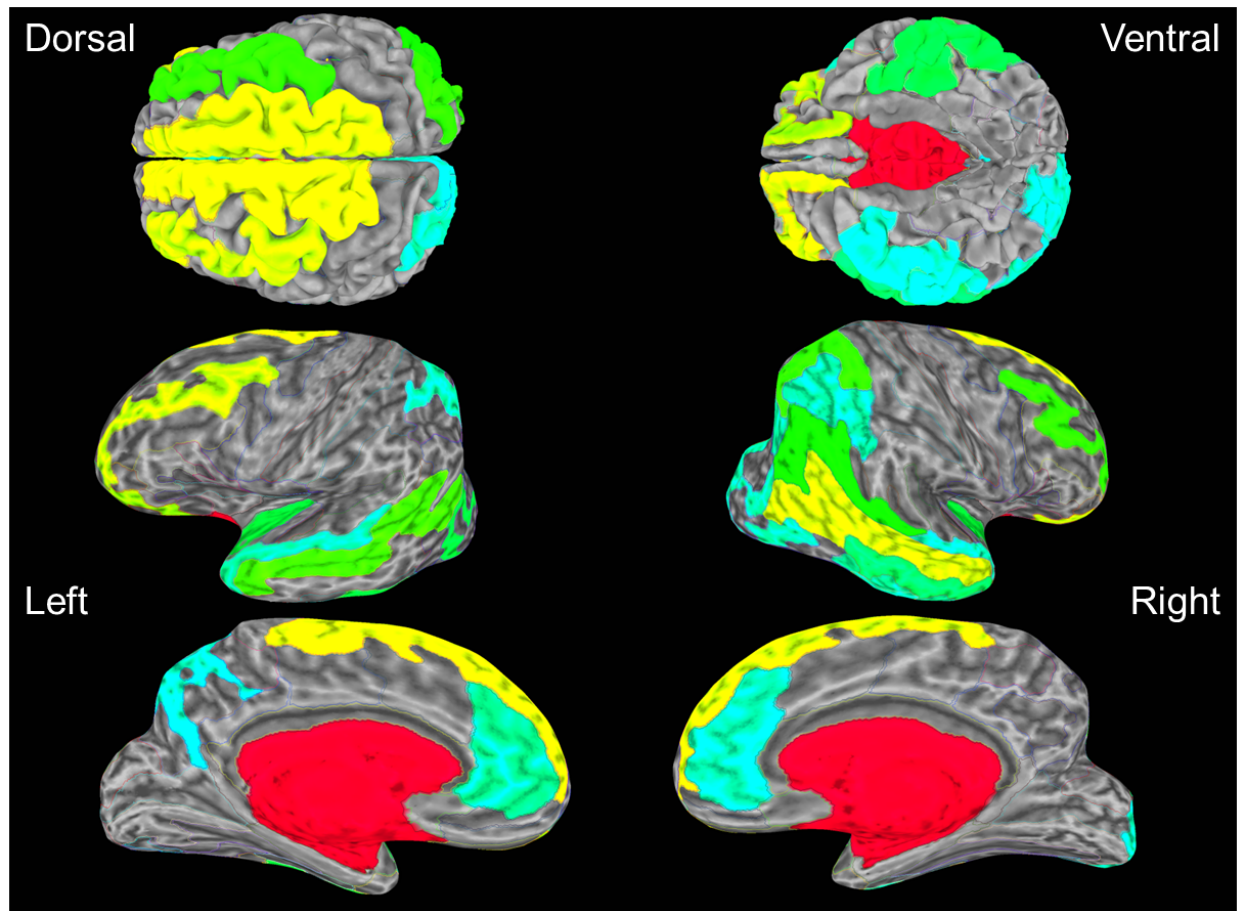


Figure 11. Cortical areas in the network of MOD (most-often-detected) connections. Blue-to-red color coding indicates the number of MOD connections projecting to each area; warmer colors indicate more connections. Images are thresholded to mask out areas with 20 or fewer connections.

Table 3. Left- and right-hemisphere areas with highest degree of MOD connections.

Left hemisphere	Degree	Right hemisphere	Degree
Medial wall	62	Medial wall	67
Superior frontal gyrus	43	Superior temporal sulcus	46
Orbital gyrus	40	Superior frontal gyrus	45
Middle frontal gyrus	40	Orbital gyrus	44
Superior temporal sulcus	33	Superior parietal lobule	36
Inferior insula/circular sulcus	29	Middle frontal gyrus	40
Inferior temporal gyrus	29	Angular gyrus	33
Lateral superior temporal gyrus	26	Inferior insula/circular sulcus	32
Anterior cingulate gyrus/sulcus	24	Middle temporal gyrus	29
Superior parietal lobule	23	Inferior temporal sulcus	26

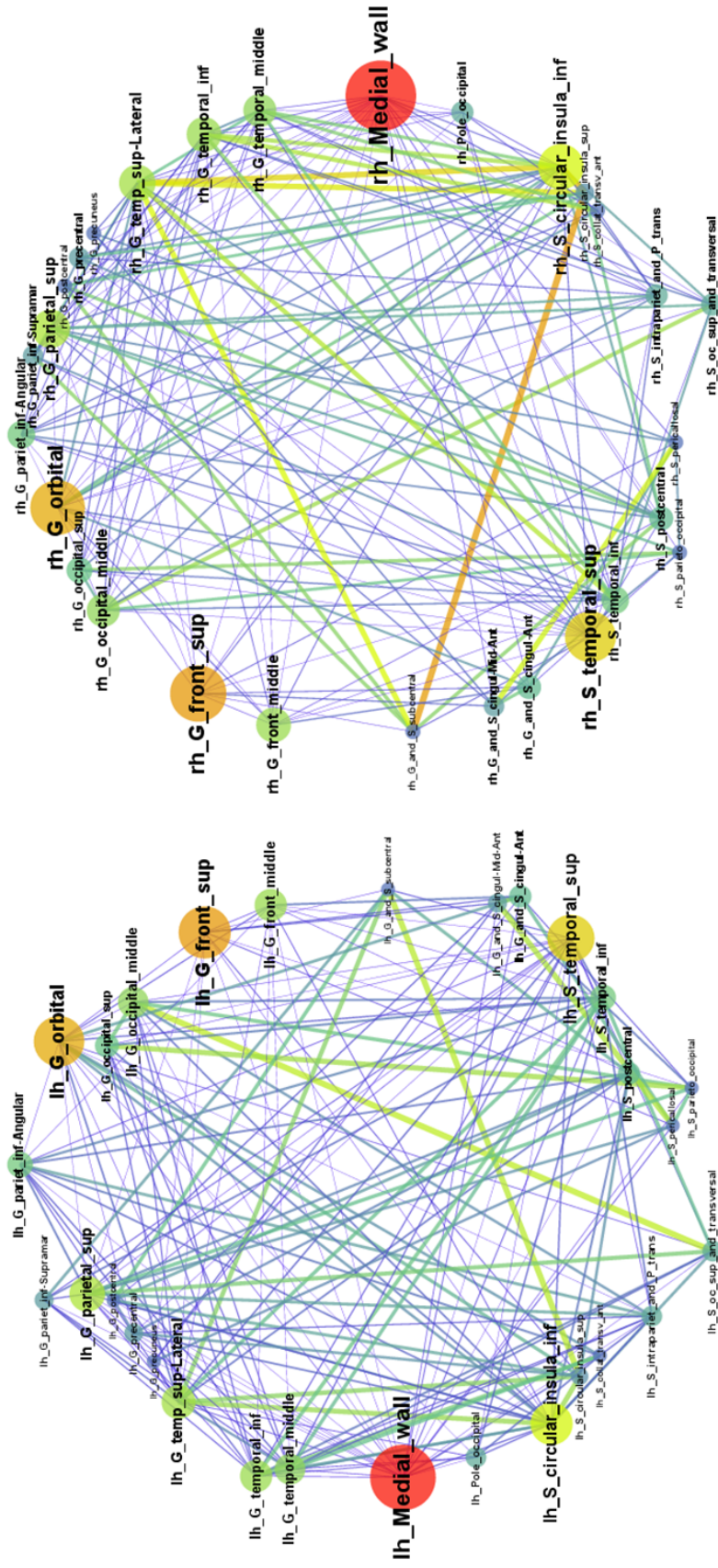


Figure 12. Graph of symmetric intrahemispheric MOD connections. Nodes with larger degrees (i.e., more connections within the plotted subset) are indicated by larger size and warmer colors, on a blue-to-red spectrum. Denser connections are indicated by thicker edges and by warmer edge colors.

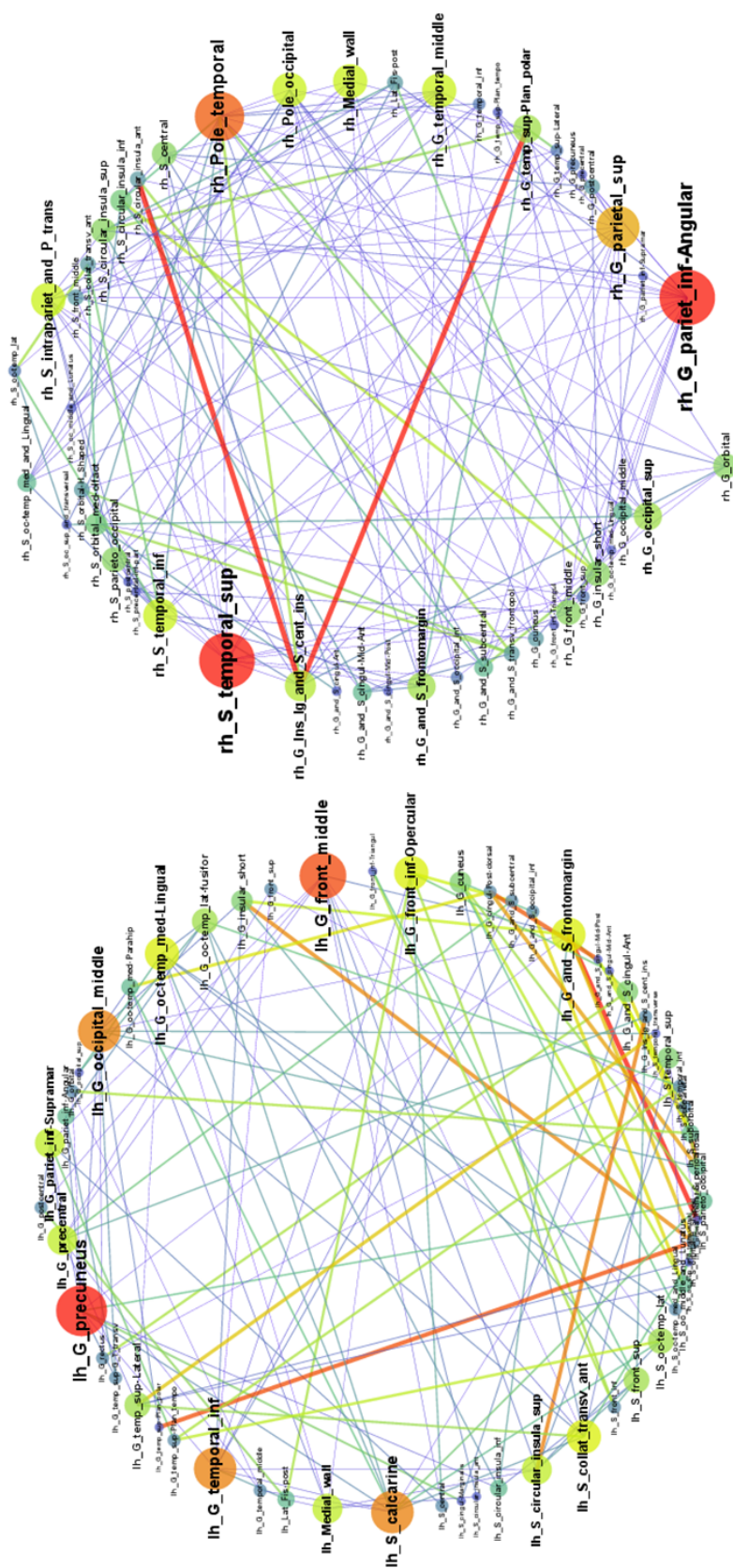


Figure 13. Graph of asymmetric intrahemispheric MOD connections. Nodes with larger degrees (i.e., more connections within the plotted subset) are indicated by larger size and warmer colors, on a blue-to-red spectrum. Denser connections are indicated by

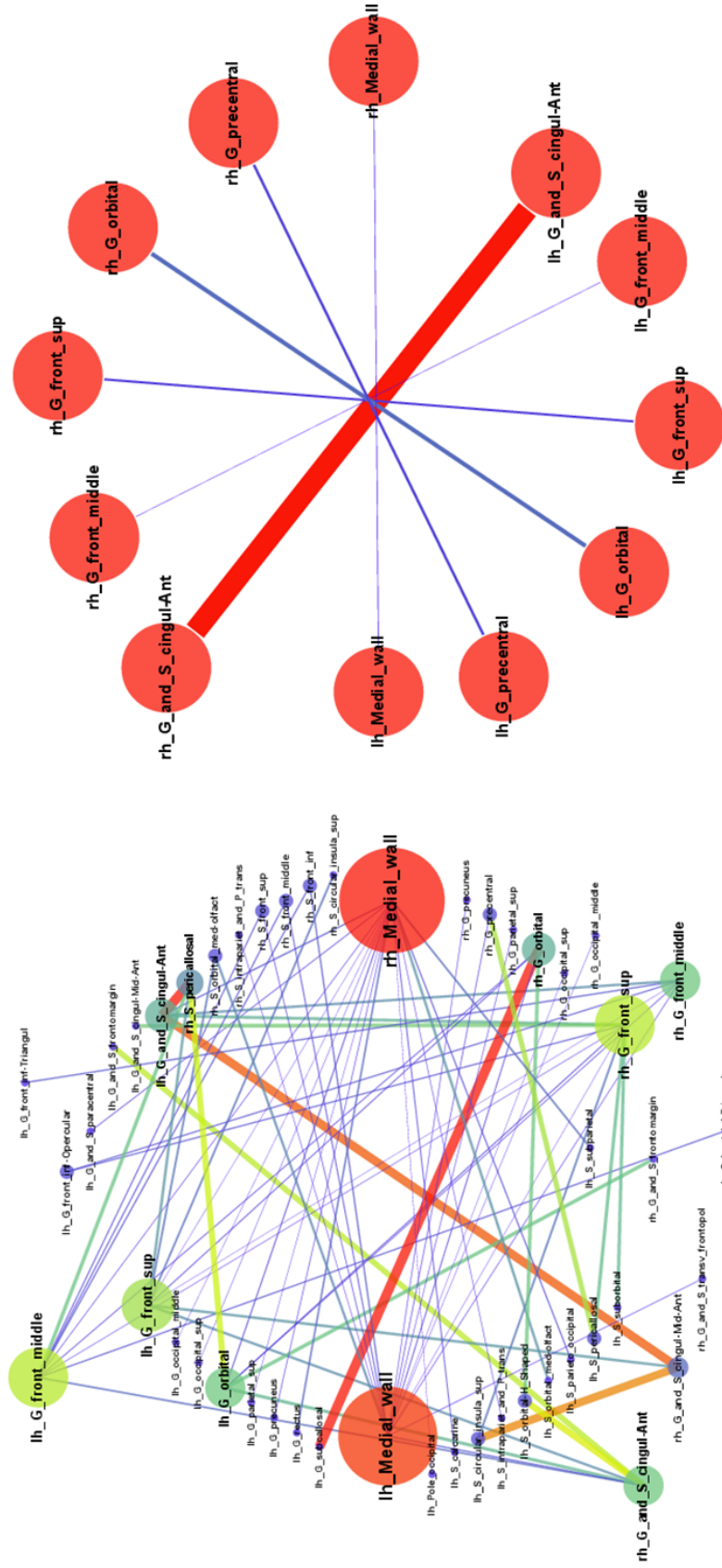


Figure 14. Graph of interhemispheric MOD connections. The left plot illustrates heterotopic connections: fibers that travel from one cortical area to a non-homologous area in the opposite hemisphere. The right plot illustrates homotopic connections, between homologous anatomical areas in the left and right hemisphere. Edge thicknesses in the homotopic plot are exaggerated relative to other plots in order to increase visibility. As above, nodes with larger degrees (i.e., more connections in the heterotopic and homotopic subsets) are indicated by larger size and warmer colors, on a blue-to-red spectrum. Denser connections are indicated by

Finally, we characterized MOD connections in terms of average fiber length. Figure 15 shows the distribution of lengths across the set of MOD connections. The mean MOD tract length was 69 mm, with a median of 63 mm (s.d.=27). We used a length criterion of 100 mm in order to identify long-distance projection tracts within the set of MOD connections. This criterion isolated a total of 116 connections, linking a total of 43 ROIs. Collapsing across hemisphere, the ten most prominent ROIs in this sub-network included the orbital gyrus (29 long-distance connections), middle frontal gyrus (26), angular gyrus (12), superior frontal gyrus (11), lateral superior temporal gyrus (10), middle occipital gyrus (10), frontomarginal gyrus and sulcus (10), precuneus (9), superior parietal lobe (9), and superior temporal sulcus (8). Several of these ROIs also rank highly in terms of total degree, including both long and short connections (Table 3). Table 4, in turn, gives the ten highest-degree areas for each hemisphere within this subset of long-distance MOD connections.

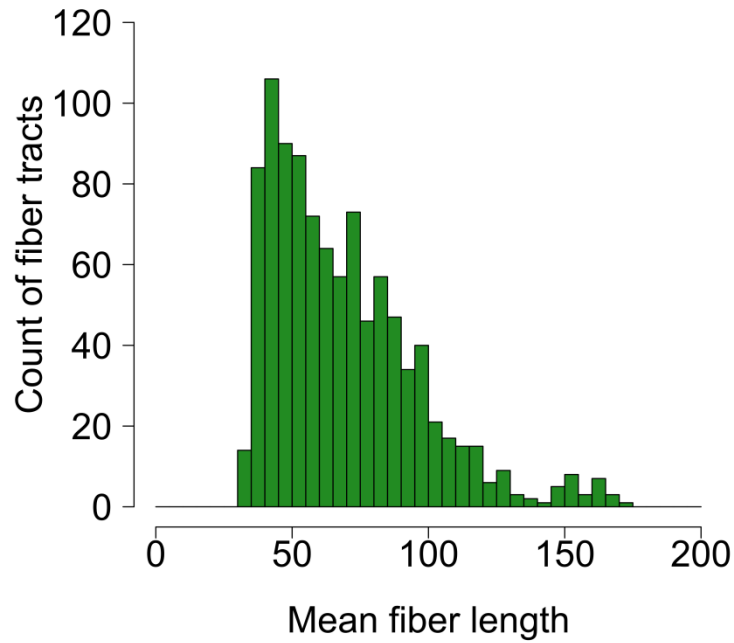


Figure 15. Distribution of average fiber lengths in MOD connection set.

Table 4. Brain areas of highest degree in long-distance subset of MOD connection network.

Left hemisphere	Degree	Right hemisphere	Degree
Orbital gyrus	14	Orbital gyrus	15
Middle frontal gyrus	12	Middle frontal gyrus	14
Lateral superior temporal gyrus	8	Angular gyrus	9
Precuneus	6	Middle temporal gyrus	8
Middle occipital gyrus	6	Superior occipital gyrus	7
Interior temporal gyrus	5	Frontomarginal gyrus/sulcus	7
Superior frontal gyrus	5	Occipital pole	6
Occipital pole	4	Superior parietal lobule	6
Superior temporal sulcus	3	Superior frontal gyrus	6
Superior parietal lobule	3	Superior temporal sulcus	5

1.3.2 Targeted Tractography Results

Targeted tracking using minimal ROI constraints enabled us to detect the corticospinal tract, arcuate fasciculus, and callosal projection fibers in all datasets (Figures 16-18). Tracking results were broadly consistent across sessions and participants, but important differences in the tractography results were also present. For example, tracking of corticospinal fibers successfully detected known projections in premotor, primary motor, and somatosensory areas (Verstynen et al., 2011). However, in a subset of datasets, erroneous projections to occipital and anterior prefrontal areas were also detected. These fibers may represent false continuations, linking corticospinal diffusion pathways in the brainstem with occipitopontine and frontopontine pathways. In almost all datasets, tractography correctly detected both the primary corticospinal and alternative motor fibers, which extend dorsally or posteriorly in the brainstem toward the cerebellum. However, for one dataset (Figure 16, subject 6, session 1), the differentiation between these structures is unclear.

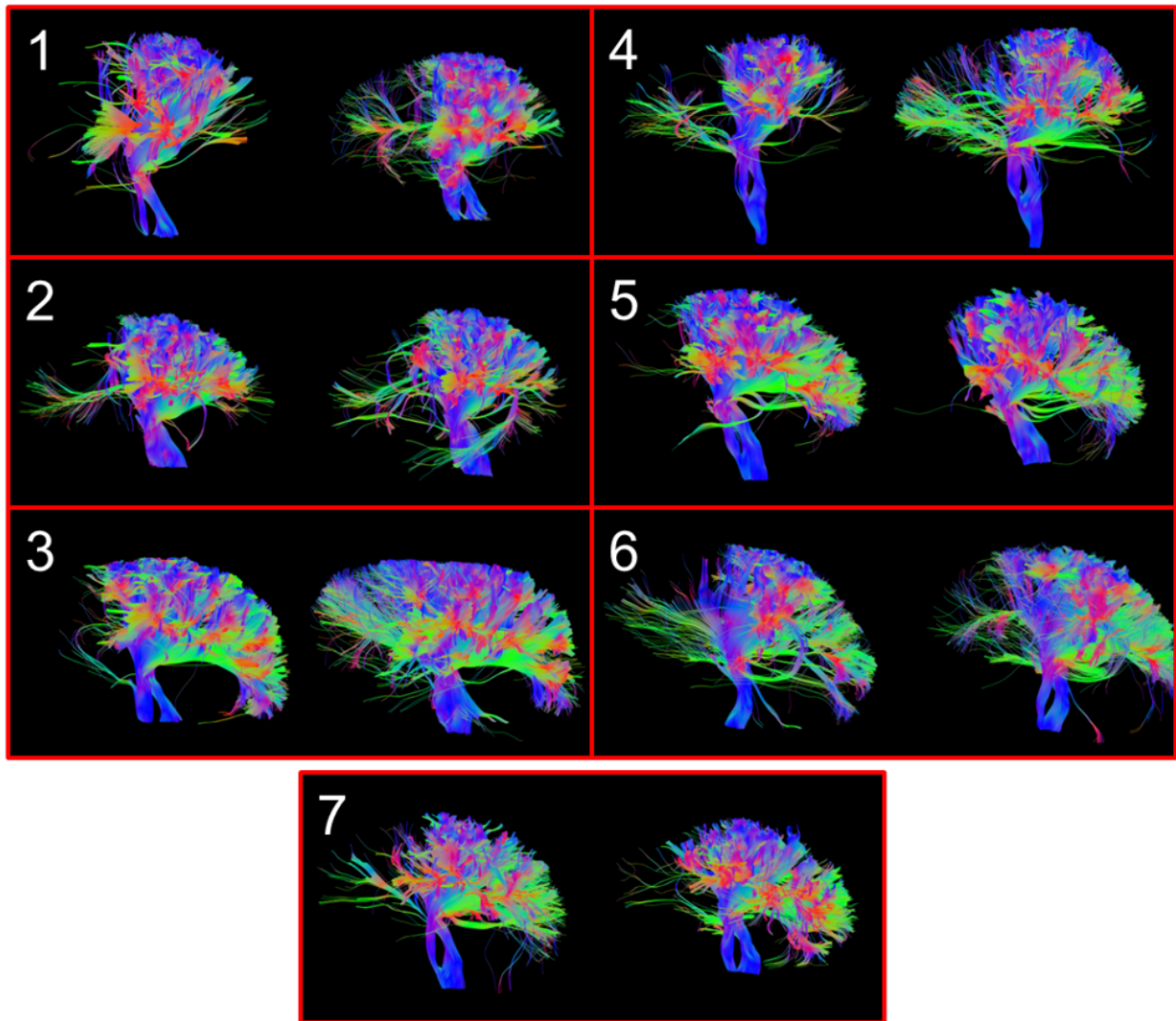


Figure 16. Targeted tracking of left hemisphere corticospinal tract. White numbers indicate participant number. Session 1 tracts appear on the left side of each panel, session 2 on the right.

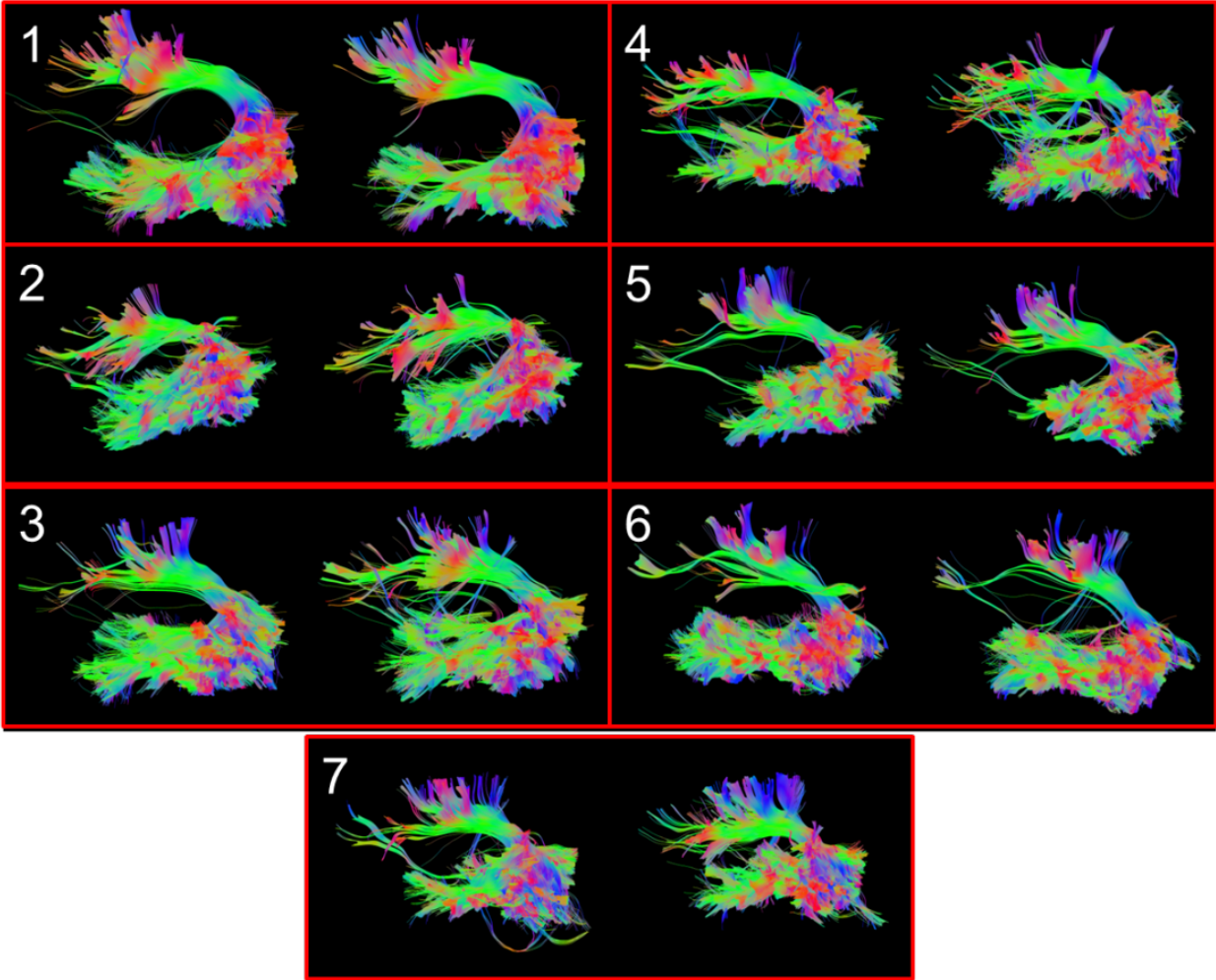


Figure 17. Targeted tracking of the left arcuate fasciculus. Layout and numbers are as in Figure 16.

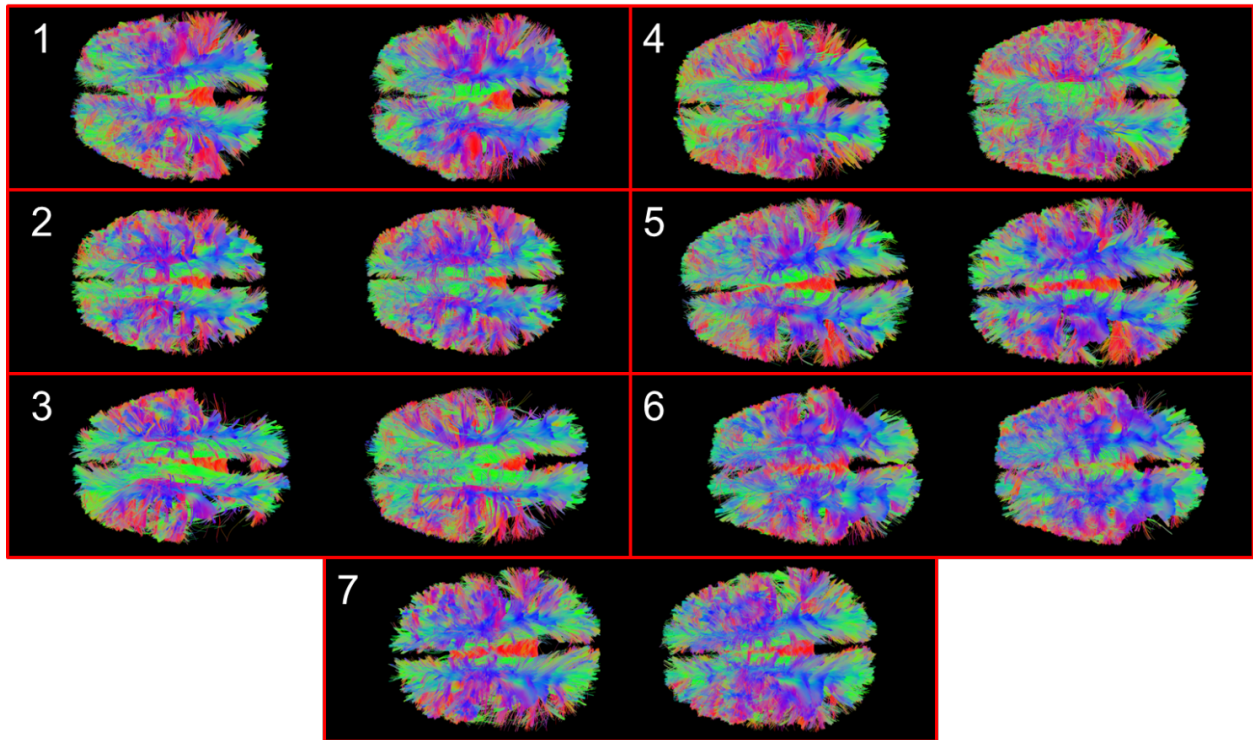


Figure 18. Targeted tracking of interhemispheric callosal projection fibers. Layout and numbers are as in Figure 16.

Tractography results for the arcuate fasciculus also broadly reproduced neuroanatomical expectations, with some visible noise. In all datasets, targeted tractography captured the expected morphology of the tract, bending around the temporo-parietal junction (TPJ) and fanning widely to terminate in the temporal lobe and prefrontal cortex. However, some discrepancies and clear errors are visible across datasets. For example, results from both sessions in participant 5 (Figure 17) show a slender fiber projection from the temporal lobes anterior to prefrontal cortex, apparently passing through the insula; this result is likely to be a false conflation of arcuate fibers with fibers of the inferior occipitofrontal fasciculus, which passes through the limen insulae into the prefrontal cortex (Nieuwenhuys, p. 593). Additionally, in some participants, the density of prefrontal terminations appears to differ considerably between sessions (e.g., participants 4 and 7). Between-session discrepancies for a given participant are not likely to result from differential coverage by ROI tractography masks, as between-session alignment was generally very good (Figure 31), and visual inspection indicated that session 1 and session 2 masks provided ample coverage in thresholded QA0 maps.

Relative to the arcuate fasciculus and corticospinal tracts, long-distance callosal projection fibers projected to a much larger brain volume and involved a larger number of region-to-region connections (Figures 18, 21, and 24). Detailed analysis of these fibers through visual inspection is not feasible; however, some features of the results depicted in Figure 18 warrant comment. The most salient of these may be differences across participants in the amount of fiber projections to parietal cortex. The low detectability of parietal projection fibers has been previously acknowledged by our group (Jarbo et al., 2012). Jarbo et al. were able to successfully tract callosal projections to posterior temporal and parietal cortex through additional targeted

tracking, which allowed only fibers with at least one endpoint in these areas; Jarbo et al. found it necessary in some cases to lower the anisotropy threshold in order to visualize these pathways.

1.3.3 Global Tractography Results

We identified fibers of interest from whole-brain tracking files using the same ROIs employed for targeted tracking. Tables 5–7 list the number of fibers thus identified for each participant and tract of interest. Notably, fiber counts differ between sessions and across participants, a result which *prima facie* could reflect either randomness in tractography seeding or between-session differences in the underlying diffusion signal. The counts of fibers per tract are considerably lower than the 100,000 fibers/tract generated in targeted tracking: an average of 5972 (std. dev. = 3491) corticospinal fibers, 1303 (542) arcuate fibers, and 9931 (3174) callosal fibers were detected. Due to the between-session discrepancies in fiber counts, we predicted greater reliability in targeted than global tracking results.

Table 5. Left corticospinal tract: fiber counts from global tractography. Number of fibers per tract was extracted from whole-brain tracking file for comparison with targeted tractography.

Participant	Session 1 fiber count	Session 2 fiber count
1	1284	5588
2	10751	9841
3	11278	11269
4	4487	4156
5	4849	5289
6	1112	6022
7	2764	4914

Table 6. Left arcuate fasciculus: fiber counts from global tractography. Number of fibers per tract was extracted from whole-brain tracking file for comparison with targeted tractography.

Participant	Session 1 fiber count	Session 2 fiber count
1	1850	1912
2	737	778
3	2163	1294
4	1279	397
5	1657	1932
6	1180	1391
7	888	783

Table 7. Callosal projection fibers: fiber counts from global tractography. Number of fibers per tract was extracted from whole-brain tracking file for comparison with targeted tractography.

Participant	Session 1 fiber count	Session 2 fiber count
1	14651	13972
2	12571	11534
3	4464	4210
4	8823	7852
5	9938	11600
6	10095	8834
7	7959	12527

1.3.4 Targeted vs. Global Tracking Comparison: Region-to-Region Connectivity

First, we assessed the degree of overlap in region-to-region connection matrices for targeted vs. global tracking (Figure 19). This analysis did not consider the density values of specific connections; rather, it only considered whether a connection was present or absent. As described in the Methods section, we evaluated between-session overlap as the average likelihood that a non-zero connection in one session would be non-zero in the same participant's second session. We investigated effects of tracking condition (global vs. targeted) and tract (corticospinal, arcuate, or callosal) upon this overlap metric using a two-way analysis of variance (ANOVA).

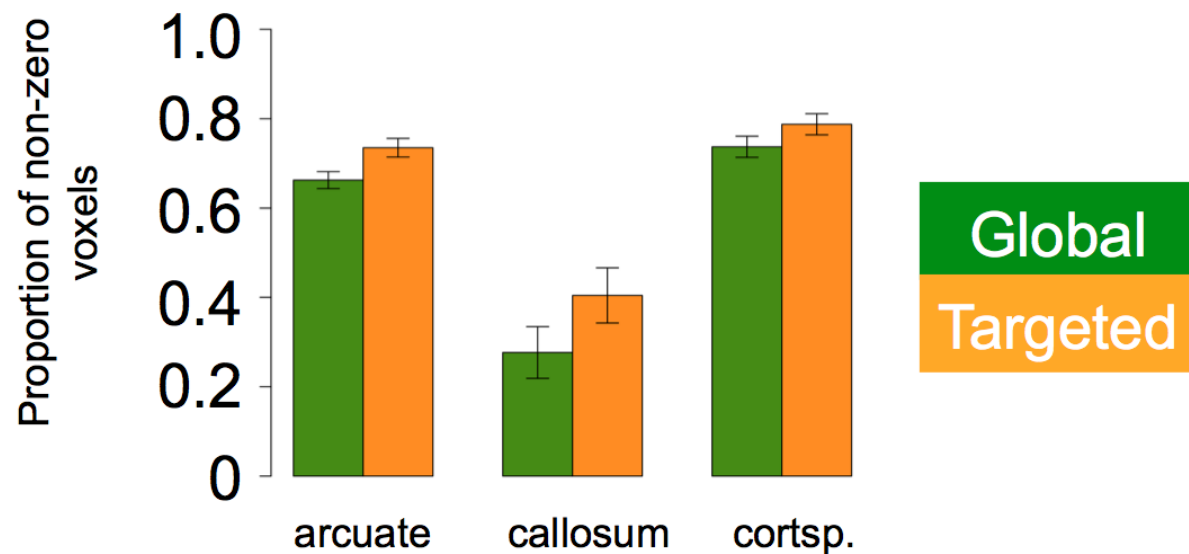


Figure 19. Between-session overlap for global and targeted tracking. Overlap indicates the average likelihood that a non-zero cell of the connection matrix in one session will also be non-zero in the second. Error bars indicate standard error of the mean.

Results indicated both a main effect of tracking condition [$F(1,30)=0.9$, $p<0.01$] and a main effect of tract [$F(2,30)=97.8$, $p<0.0001$]. The interaction of condition and tract was non-significant [$F(2,30)=0.8$]. This pattern of results appeared to indicate generally higher between-session repeatability of connections in targeted than. global tracking (Figure 19), an interpretation which was corroborated by a paired, two-tailed t-test [$t(20)=6.0$, $p<0.0001$]. The mean difference between targeted and global overlap scores in this test was 0.08, which constituted a substantial increase in the likelihood of replicating specific connections across sessions. In tract-specific comparisons, overlap of arcuate fasciculus connections averaged 0.74 (0.06) for targeted tracking, compared to 0.66 (0.05) for global tracking, and this difference was statistically reliable [$t(6)=3.7$, $p<0.01$]. The corresponding comparison for the corticospinal tract was non-significant [$t(6)=1.7$, $p<0.15$], although mean overlap was similarly higher for targeted vs. global tracking results: 0.79 (0.06) vs. 0.74 (0.06), respectively. Overlap of callosal projections averaged 0.40 (0.16) for targeted tracking and 0.28 (0.15) for global tracking; this effect was significant [$t(6)=8.9$, $p<0.0005$]. Across tracking conditions, overlap was lower for callosal connections, relative to those of the AF [$t(13)=8.5$, $p<0.0001$] and the corticospinal tract [$t(13)=10.8$, $p<0.0001$]. Overlap of corticospinal connections was higher than those of the AF [$t(13)=4.0$, $p<0.005$].

As in the intercortical connectivity analysis, we further assessed reliability by examining the between-session correlation of fiber density values for all detected connections. Per expectations, correlations were considerably reduced by “one-off” connections detected in the overlap analysis. In order to determine the reliability of density values for connections repeated across sessions, we performed a follow-up analysis, restricting our focus to connections detected in

common across each participant's two sessions. We report only correlations from this focused analysis (Figure 20), since it communicates information complementary to the overlap analysis.

We assessed effects of tracking condition on fiber density reliability by testing the significance of the difference between paired correlation coefficients. This was done by performing a Fisher's r-to-z transformation (Kenny, 1987) on each coefficient, then testing it against the coefficient derived from the opposing tracking condition in the same fiber tract and participant. The test statistic, Z , is given for individual cases in Table 8; significance was based on a Bonferroni-corrected familywise error rate of 0.05. It is important to note that the p-values in Table 8 are based on a one-tailed test, corresponding to our hypothesis of greater reliability for targeted than global tracking. Results indicated no consistent advantage in fiber density correlations for targeted tracking: across the 3 tracts and 7 participants tested, only one observation was consistent with that hypothesis. In fact, visual inspection of the strip plot in Figure 20 suggests that the opposite trend was more prevalent; however, a test for greater correlations in the global condition was significant in only 1 of 7 participants for the AF; 2 of 7 for callosal fibers; and 2 of 7 for the corticospinal tract. Thus, while the between-session overlap of binarized connection matrices did differ across tracking conditions, correlations of fiber density did not.

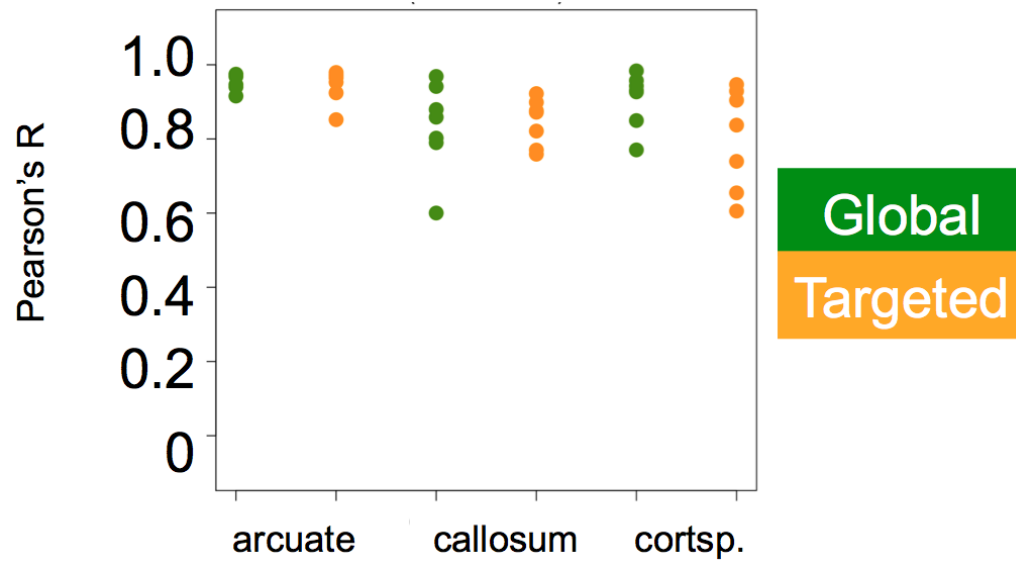


Figure 20. Between-session reliability of fiber density with global and targeted tracking.

Values are based on intercortical connections that were non-zero in both sessions of a participant's tractography results. Points represent the Pearson's correlation coefficient (r) for a single participant, plotted separately for each tract of interest and tracking condition. All correlations were statistically significant.

Table 8. Difference between correlation coefficients from targeted vs. global tracking. The test statistic represents the hypothesis that between-session correlation of fiber density was higher under targeted than global tracking conditions. P-values reflect a one-sided test, following this hypothesis. The asterisk indicates a single significant result, after correction for multiple comparisons.

Arcuate fasciculus		Z (Targeted – global)	p
	s1	-1.1	0.9
	s2	-2.5	1
	s3	-2.9	1
	s4	-0.5	0.7
	s5	2.4	0.01
	s6	2.8	0.003
	s7	0.1	0.5
Callosal fibers			
	s1	2.0	0.02
	s2	-0.4	0.7
	s3	2.2	0.01
	s4	-2.2	1
	s5	-3.8	1
	s6	-7.0	1
	s7	3.6	0.0001*
Corticospinal tract			
	s1	-1.6	1
	s2	-1.1	0.9
	s3	-0.3	0.6
	s4	-3.6	1
	s5	-1.2	0.9
	s6	-4.4	1
	s7	0.7	0.2

Finally, we assessed the power of global vs. targeted tractography in terms of the number of region-to-region connections detected (Figure 21). A connection was counted for a given participant if it was non-zero in both DSI sessions. A two-way ANOVA with random effect of participant indicated a main effect of tracking condition [$F(1,30)=20.1$, $p<0.0001$], a main effect of tract [$F(2,30)=27.5$, $p<0.0001$], and an interaction of tract with condition [$F(2,30)=12.4$, $p<0.0001$]. The effect of tracking condition reflected a greater number of connections detected across sessions in targeted than global tracking [$t(20)=3.5$, $p<0.005$]. The main effect of tract partially reflects between-tract differences in size and breadth of projection, which are incidental to the current analysis. However, the interaction of tracking condition with tract warrants comment, as it reflects a between-tract difference in the gain in detection power associated with targeted tracking. For the arcuate fasciculus and corticospinal tract, targeted tracking produced 1.99 and 1.75 times as many fibers as global tracking, respectively. For the callosal tract, this ratio was 4.13, indicating a much larger gain in the number of intercortical connections detected with targeted tracking.

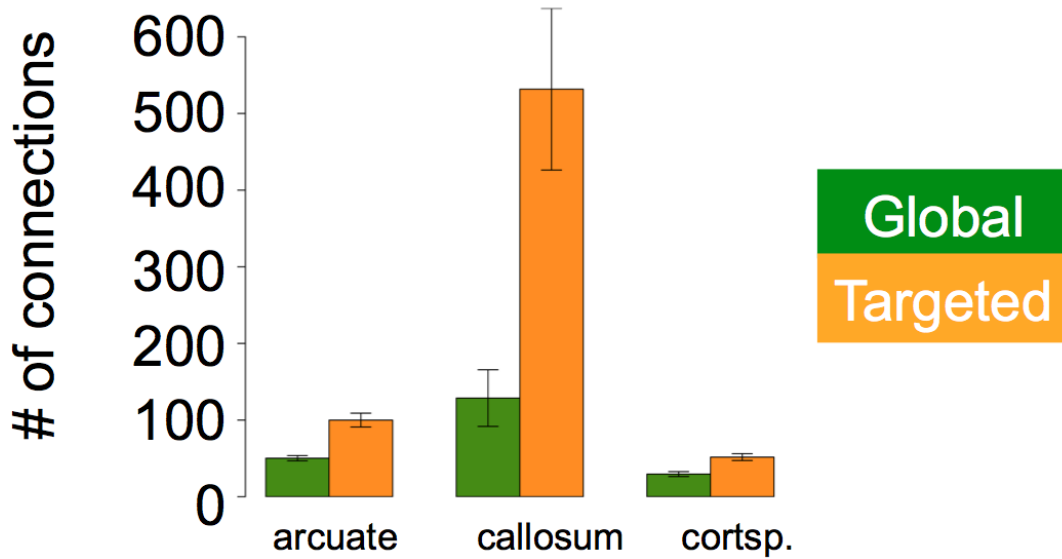


Figure 21. Comparison of detection power under global vs. targeted tracking conditions.

Detection power was assessed by the number of connections which were non-zero in both of a participant's DSI datasets.

1.3.5 Targeted vs. Global Tracking Comparison: Voxelwise Analysis

While the above results compare global and targeted tracking between anatomically-defined ROIs, we also wished to examine reliability and detection power on the level of a single voxel. We hypothesized that targeted tracking would be associated with greater between-session reliability, both in the spatial overlap of tracts and the voxelwise correlation of fiber counts. We further predicted that targeted tracking would be associated with greater detection power, resulting in fuller termination fields and tract bodies than in global tracking. We approached this question in two ways, by comparing both full-tract and endpoint-only density maps derived from

global and local tracking. In order to address the problem of residual spatial error after alignment of fiber tracts from different sessions, we also investigated the effect of spatial smoothing upon these results. Smoothing values between 2 and 20 mm FWHM were implemented; however, we report only analyses based on a smoothing kernel of 6 mm FWHM. This restriction was made, in part, because voxelwise correlation of participants' unweighted DSI images (i.e., "b0" images), which hold richer information about anatomical structure than fiber density maps, indicated maximal similarity between each participant's two sessions with smoothing of 6 mm FWHM (Figure 31). Furthermore, this value is the most anatomically plausible in the range of smoothing values investigated, since it represents approximately 1-2 voxels' worth of residual spatial error. Spatial misalignment following affine registration was not expected to exceed this magnitude, so we reasoned that smoothing in excess of 6 mm would induce spurious similarity between density maps. Indeed, spatial overlap between density maps and the pairwise correlations of voxelwise fiber counts increased monotonically over from 2 to 20 mm FWHM.

Table 9. Descriptive statistics for endpoint maps from global vs. targeted tracking. Results represent an average across tracts. Numbers in parentheses indicate standard deviations in the between-subject mean.

Tracking condition	Endpoint Overlap	Density correlation	Number of voxels
Global	0.76 (0.12)	0.39 (0.28)	48548 (39218)
Targeted	0.80 (0.12)	0.39 (0.27)	63578 (47503)

Table 10. Descriptive statistics for full-tract density maps from global vs. targeted tracking.

Results represent an average across tracts. Numbers in parentheses indicate standard deviations in the between-subject mean.

Tracking condition	Full-tract Overlap	Density correlation	Number of voxels
Global	0.81 (0.10)	0.62 (0.30)	58361 (39372)
Targeted	0.83 (0.10)	0.60 (0.30)	70842 (44874)

We first report the results of endpoint map analysis. As in region-to-region analysis, we conducted an ANOVA on overlap scores, with fixed effects of condition and tract, plus a random effect of participant. Consistent with hypotheses, we found a modest effect of tracking condition upon endpoint map overlap [$F(1,30)=6.3$, $p<0.05$]. Mean overlap in the global condition, collapsing across tract, was 0.76 (0.12), compared to 0.80 (0.12) in the targeted condition (Table 9), representing a 4% increment. This difference was statistically reliable [$t(20)=5.0$, $p<0.0001$], confirming that targeted tracking was associated with higher between-session overlap of endpoint fields. The ANOVA also produced a robust main effect of tract [$F(2,30)=66.5$, $p<0.0001$]. Post-hoc tests indicated superior overlap in callosal endpoint maps, relative to the AF [$t(13)=7.6$, $p<0.0001$] and the corticospinal tract [$t(13)=12.2$, $p<0.0001$]. However, this result must be considered in terms of the much lower overlap in region-to-region connection matrices for callosal fibers (see Figure 19, left-hand plot): high between-session overlap in callosal projection fields may not actually indicate reliable fiber connectivity, but rather be a trivial consequence of its wide fanning throughout frontal, parietal, and posterior temporal cortices.

Overlap scores for corticospinal and arcuate fibers were statistically indistinguishable [$t(13)=0.0$, $p<1$].

Between-session correlation of voxelwise fiber counts (Table 9 and Figure 22, left-hand plot) was generally lower than the correlation of region-to-region connection densities and was highly variable. This difference was expected due to the much higher granularity of voxelwise analysis, which entails greater susceptibility to small spatial misalignments or to small deviations in fiber trajectory. As in the region-to-region analysis, we performed pairwise tests of the difference between correlation coefficients from targeted vs. global tracking. Results suggested no consistent difference in correlations between these conditions. For the AF, 4 of 7 participants exhibited improved correlations in the targeted tracking; however, only 2 of 7 participants showed this pattern for callosal fibers, and just 1 of 7 for the corticospinal tract. Across all 21 comparisons, the test statistic Z had the hypothesized sign in 11 cases, and the opposite result in 12 cases. Thus, as in ROI-based analysis of our three tracts of interest, the between-session correlation of fiber density values did not appear to differ by tracking condition.

Detection power (Figure 24) did differ by tracking condition [$F(1,30)=50.6$, $p<0.001$], indicating significantly more voxels detected in tract termination fields for targeted vs. global tracking [$t(20)=6.5$, $p<0.0001$]. Collapsing across tract, global tracking produced termination fields which averaged 48548 voxels (39218), compared to 63578 (47503) for targeted tracking. The main effect of tract was also significant [$F(2,30)=780.7$, $p<0.0001$]. This effect reflected the different morphology and density of the three tracts tested. In particular, callosal projection fibers had significantly larger termination fields than the AF [$t(13)=25.9$, $p<0.0001$] and corticospinal tract [$t(13)=21.2$, $p<0.0001$]. Corticospinal termination fields were also larger than

arcuate termination fields [$t(13)=8.0$, $p<0.00001$]. The interaction of tract with condition was also significant [$F(2,30)=7.9$, $p<0.005$].

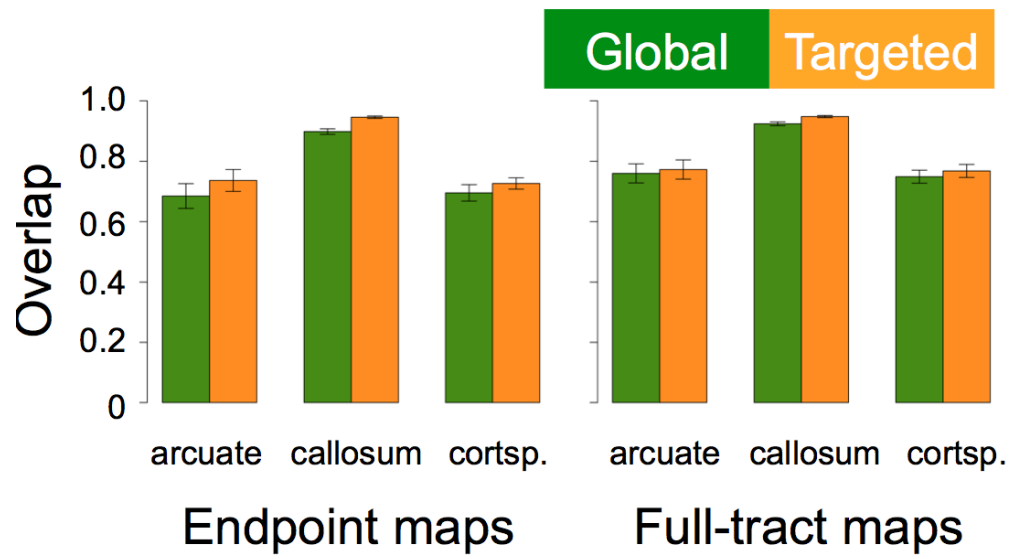


Figure 22. Voxelwise between-session overlap for global and targeted tracking. Results are plotted separately for each tract of interest and tracking condition.

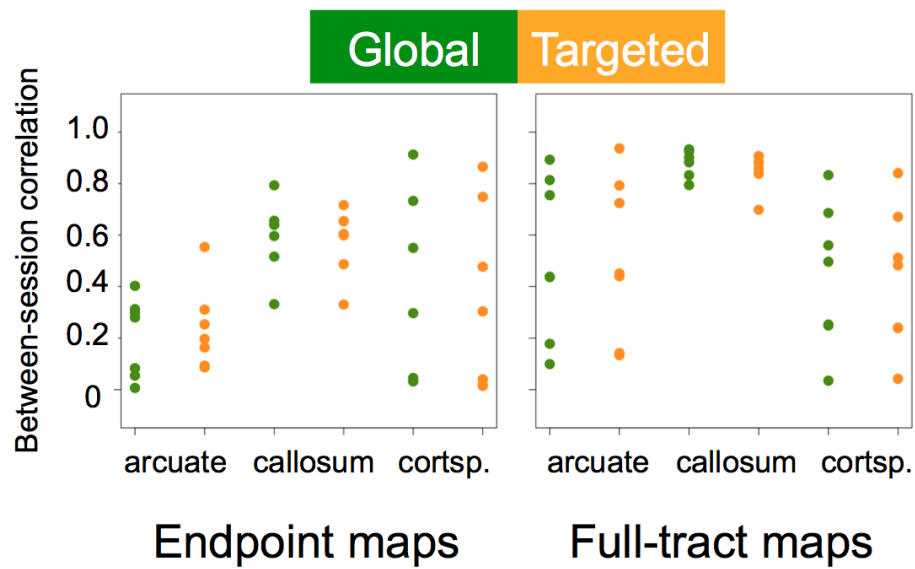


Figure 23. Voxelwise between-session density correlations for global and targeted tracking.

Results were calculated over the set of voxels that were non-zero in both sessions of a participant's tractography results (see Figure 12). The left plot shows correlations in tract termination fields, while the right shows correlations of full-tract density maps. Each point represents the Pearson's correlation coefficient, r , for a single combination of participant, tract, and tracking condition.

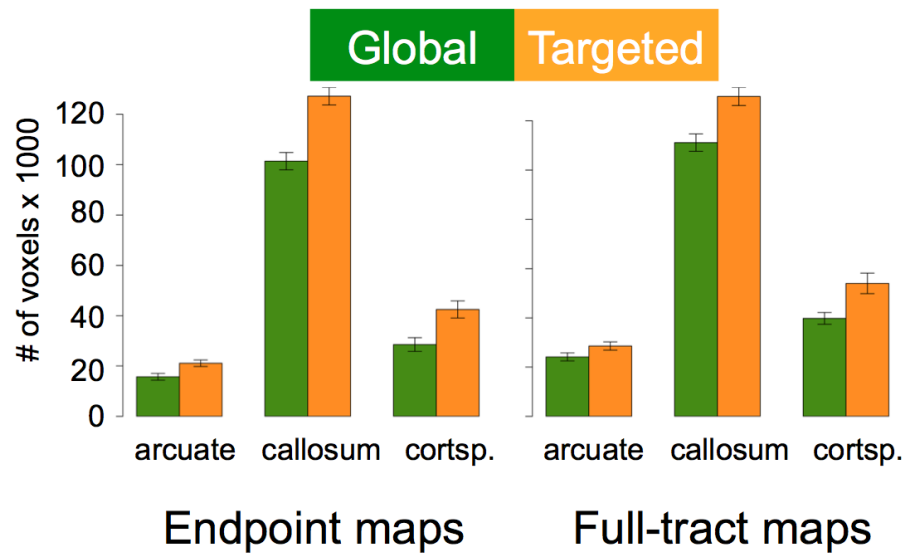


Figure 24. Voxelwise detection power for global and targeted tracking. Detection power was measured by the count of overlapping voxels in density maps from a participant's two tractography datasets. The left plot is based on tract termination fields, while the right plot is based on full-tract density maps.

Next we analyzed overlap, voxelwise correlation, and detection power in full-tract density maps (Figures 22-24, right-hand plots). In contrast to intercortical connection matrices or endpoint maps, full-tract maps convey information about the entire body of a tract and can be viewed as a rough summary of tract position and morphology. In contrast to endpoint maps, voxelwise overlap did not differ by tracking condition [$F(1,30)=1.9$, $p<0.2$]. The main effect of tract was significant [$F(2,30)=74.3$, $p<0.0001$]; as in the endpoint map analysis, overlap was significantly greater for callosal fibers than for the AF [$t(13)=8.2$, $p<0.0001$] and corticospinal fibers [$t(13)=13.0$, $p<0.0001$]. Between-session overlap scores for the AF vs. corticospinal tract

were not significantly different [$t(13)=0.5$, $p<0.7$]. Overlap in full-tract density maps was slightly higher than in endpoint-only maps (Table 10), and this difference was statistically reliable [$t(41)=7.9$, $p<0.0001$]. This difference in overlap rates is not surprising, since the bodies of fiber tracts exhibit more homogeneous density than the fanning portions near their termination fields, and thus may be less susceptible to residual spatial error; moreover, tract bodies travel in deep white matter, where anisotropy values are larger and more reliable than near the gray/white matter boundary (Farrell et al., 2007).

Tests of correlation coefficients suggested no effect of tracking condition for the AF and corticospinal tract. Individual tests for both of these tracts yielded a mixture of results consistent with and contrary to hypotheses. Individual correlation coefficients for these tracts were also highly variable across individuals. Unexpectedly, higher correlations were found for the global condition in callosal fibers; this result, which ran contrary to hypotheses, was present and robust in 6 of 7 participants. This finding implied that voxelwise fiber densities in the body of callosal tracts were consistently more reliable for global than targeted tracking. These between-tract differences, which may result from differences in tract density and morphology, are addressed further in the discussion.

Finally, a comparison of voxel counts in the between-session intersection of each participant's density maps indicated a main effect of condition [$F(1,30)=36.8$, $p<0.0001$], a main effect of tract [$F(2,30)=772.9$, $p<0.0001$], and a modest interaction of the two factors [$F(2,30)=4.2$, $p<0.05$]. This pattern of results mirrored the results of the endpoint map analysis, and indeed, appeared to be attributable to differences in the size of tract termination fields. While a comparison of voxel counts for targeted vs. global tracking was significant [$t(20)=6.6$, $p<0.0001$], both the t-statistic and the difference of the means (12481) were similar to the

corresponding analysis of endpoint maps. Taken together, these results suggest that the global and targeted tracking filled similar voxel volumes along most of the length of a fiber tract's body, but that targeted tracking produced fuller termination fields than global tracking.

1.3.6 Sources of Error in Tractography: Summary of Follow-up Analyses

1.3.6.1 Assessment of Reseeding Error

Tests of reseeding error suggested that random seeding of tractography did not influence tractography results, except at very low fiber counts. Figure 25 shows the reliability of repeated within-session tracking over a range of fiber counts, as assessed by correlating intercortical connection matrices for two runs of tractography on each participant's session 1 data. At a fiber count of 5000, all between-session correlation coefficients exceeded 0.95; with 100,000 fibers, all coefficients exceeded 0.99. For comparison, we note that 500,000 fibers were generated per dataset in the intercortical connectivity analysis. These results indicate nearly identical results between runs of the deterministic tractography algorithm when the diffusion space is well-seeded. By extension, this consistency suggests that differences in random seeding did not contribute to between-session error.

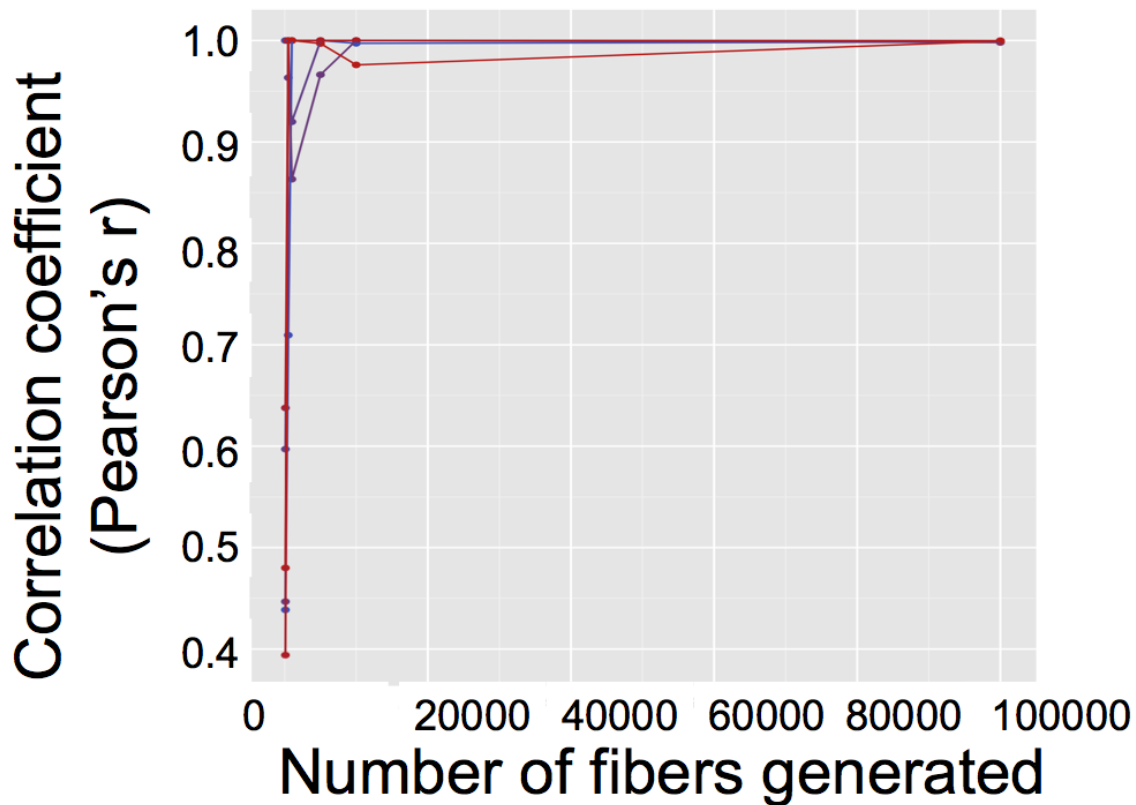


Figure 25. Assessment of reseeding error. Plot shows correlation coefficients for repeated runs of the tractography algorithm within subject. Intercortical fiber connections were tracked at fiber counts of 100, 500, 1000, 5000, 10000, and 100000. Reseeding error was assessed in one session for each of the 7 participants; each line indicates the similarity curve for a single participant.

1.3.6.2 Parameter Sensitivity Analysis

Different parameter choices and their optimality to the specific DSI dataset being analyzed represented another possible source of between-session error in tractography results. For a single dataset, we investigated differences in corticospinal tractography results across changes in three tracking parameters: tracking threshold (i.e., anisotropy threshold), smoothing (i.e., momentum), and angular threshold. We predicted that extreme parameter choices might distort tracking results, but that neuroanatomically valid fiber projections would be observable across a reasonable range of values for each of these variables. The parameter sensitivity analysis was performed in two steps. In Step 1, we eliminated parameter combinations or which produced fibers with low scores on a neuroanatomical accuracy test. Figure 16 shows the distribution of accuracy scores and fiber counts from Step 1. Of the 250 parameter combinations tested, 20% (50 combinations) failed to produce any fibers. These parameter combinations, as well as cases which yielded net accuracy scores of less than 0.5, were excluded from Step 2. Net accuracy was computed by calculating hit and false alarm rates (the proportion of ROIs in the hit and false alarm ROI lists which contained a non-zero number of fiber terminations), then subtracting the false alarm rate from the hit rate.

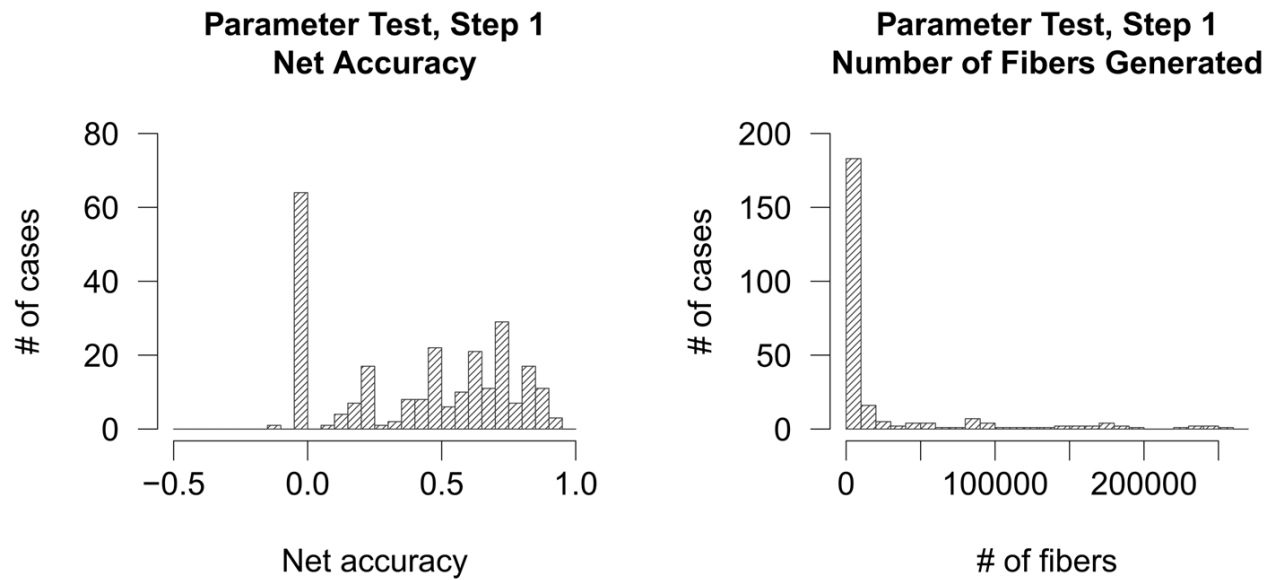


Figure 26. Results from Step 1 of the parameter sensitivity analysis. Left: net accuracy was computed for each tract as the proportion of “hit” regions to which it projected, minus the proportion of false alarm regions that it contacted. Parameter combinations that generated tracts with net accuracy < 0.5 were omitted from further analysis in Step 2. Right: in Step 1, the stopping criterion for each tracking job was the generation of 25 million random seeds, a subset of which produced virtual fibers. Parameter combinations that failed to produce any fibers (20 percent of the total count of combinations) were also omitted in Step 2. Of the 250 parameter combinations tested in Step 1, 119 were thus discarded, leaving 131 for re-test.

Of the 250 parameter combinations tested in Step 1, 131 were re-tested in Step 2. We generated a uniform count of 100,000 fibers with each of the surviving parameter combinations, then recomputed results for the neuroanatomical accuracy test. The distribution of net accuracy scores (Figure 27) indicates that tractography results are a complex, interactive function of these tractography variables. The various and combined effects of the tested tractography variables were assessed by regressing tracking threshold, angle threshold, smoothing values, and all possible interactive terms onto net accuracy scores. Parameter combinations that were not included in Step 2 were treated as missing observations. This model was highly significant overall [$F(7,123)=5.8$, $p<0.00001$] and produced an adjusted R^2 valued of 0.21. Main effects were observed for all three tractography variables: tracking threshold was associated with a regression coefficient of -0.097 [s.e. of 0.03; $t(123)=-3.3$, $p<0.005$]; angle threshold was associated with a coefficient of -0.17 [s.e. of 0.06; $t(123)=-2.8$, $p<0.01$]; and smoothing was associated with a coefficient of -0.15 [s.e.=0.05; $t(123)=-2.7$, $p<0.01$]. A coefficient of 0.03 (s.e. of 0.009) was found for the interaction of tracking and angle thresholds [$t(123)=3.5$, $p<0.001$]. Additional significant effects were found for the interaction of tracking threshold with smoothing value [coefficient of 0.03, s.e. of 0.0097; $t(123)=2.7$, $p<0.01$] and for the three-way interaction of tracking threshold, angle threshold, and smoothing [coefficient of -0.007, s.e. of 0.003; $t(123)=-2.6$, $p<0.01$]. Of the large number of model terms tested, the main effect of tracking threshold and the interaction of tracking threshold with angle threshold survived a Bonferroni correction for multiple comparisons. The main effect of angle threshold just missed significance by this procedure (corrected p-value of 0.0501, compared to an alpha level of 0.05).

Visual inspection of Figure 27 facilitates interpretation of these statistical results. Especially notable is the fact that very strict angle thresholds were associated with low accuracy: no

tractography cases using an angle threshold of 15° satisfied the inclusion criteria for Step 2. Moreover, lenient angle thresholds were only associated with high accuracy when tracking thresholds were also increased. In contrast, the combination of permissive angle thresholds and low tracking thresholds produced a large number of false continuations and improbably sinuous fibers that lowered overall accuracy considerably. This pattern of results caused the exclusion of a large number of parameter combinations from retesting in Step 2 (see the large n/a regions at angle thresholds of 60° and 75° in Figure 27). Parameter combinations with angle thresholds of 60 - 75° and low tracking thresholds (less than or equal to a QA value of 0.4) had a mean hit rate of 1.0; however, the mean false alarm rate for these cases was also high (0.62), causing a decline in net accuracy. These results suggest that when tracking threshold is eased in order to allow the detection of strongly curving fibers, some other constraint—namely, tracking threshold—must be adjusted in order to maintain control over noise in tractography results. The effects of smoothing values were more subtle and less statistically reliable. However, it is notable that with an angle threshold of 30° , high smoothing values (0.8-0.9) resulted in a large number of low-accuracy results in Step 1. Similarly, a smoothing value of 0.9 led to poor accuracy in Step 1 results at high tracking thresholds and angle thresholds of 45 - 60° . These results could indicate that when fiber trajectories are tightly controlled (e.g., by strict angle or tracking thresholds), high smoothing adds a further constraint that undesirably straightens fibers and prevents them from following naturally curving diffusion pathways.

We additionally computed a similarity matrix for all tracts produced in step 2, in order to test whether tracts that ranked in the top quartile for anatomical accuracy were also characterized by higher mutual similarity than tracts in lower quartiles. The similarity matrix was computed by converting each tract into a NifTi-format density image, transforming the voxel matrix for each

image into a long vector (length of 460800, resulting from image dimensions of 96 x 96 x 50 voxels), then finding the Pearson's correlation coefficient between each vector and every other vector. The average mutual correlation within each quartile is shown in Figure 29.

We tested between-quartile differences by applying a Fisher's r-to-z transformation to all correlation coefficients, then comparing quartiles using two-sample, two-tailed t-tests, assuming heteroscedasticity. The results indicated higher mutual similarity in the top quartile than in the first [$t(639)=2.4$, $p<0.05$] and second quartiles [$t(504)=6.4$, $p<0.0001$]. These results suggested convergence of tracking results with increasing stringency of the anatomical accuracy criterion. However, tracts in the top quartile were less similar on average than those in the third quartile [$t(722)=-3.6$, $p<0.0005$], possibly indicating a cluster of similar tracking results with lesser overall accuracy.

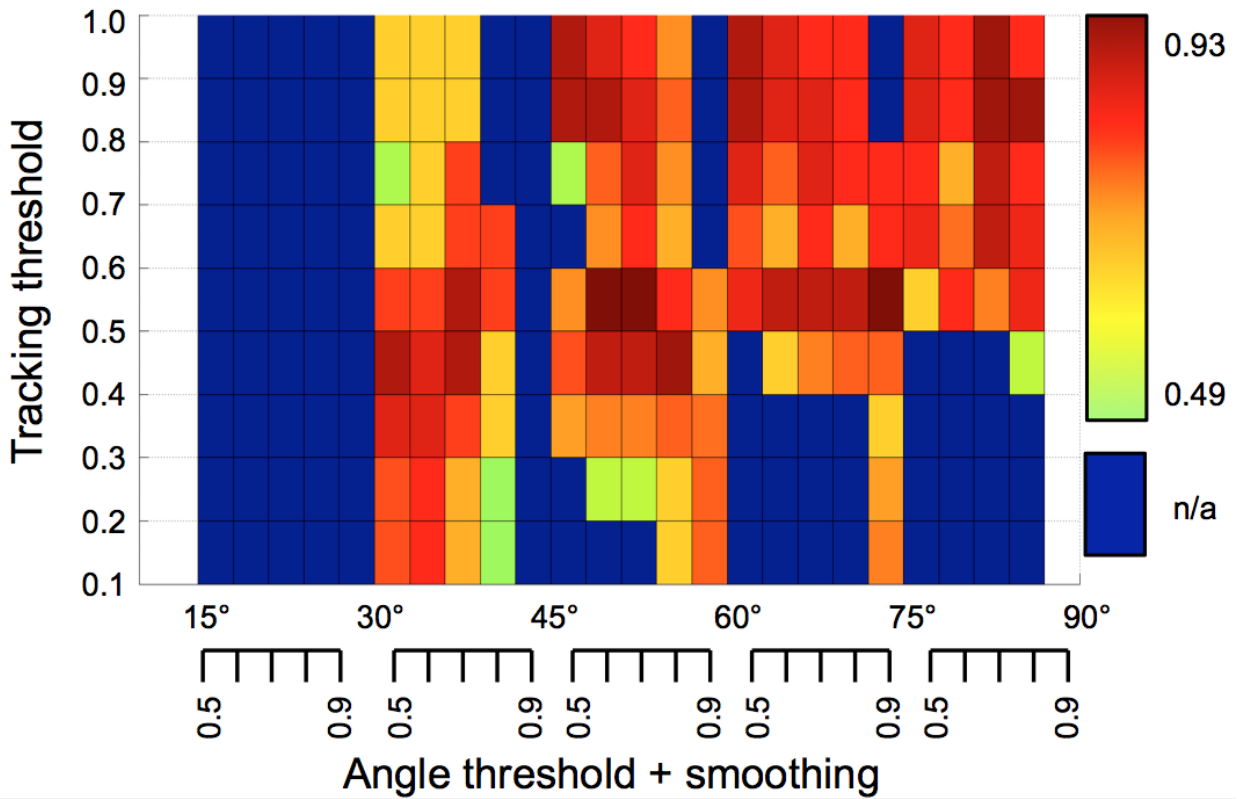


Figure 27. Results from Step 2 of the parameter sensitivity test. A total of 131 parameter combinations were tested. Net accuracy was computed as hit rate – false alarm rate.

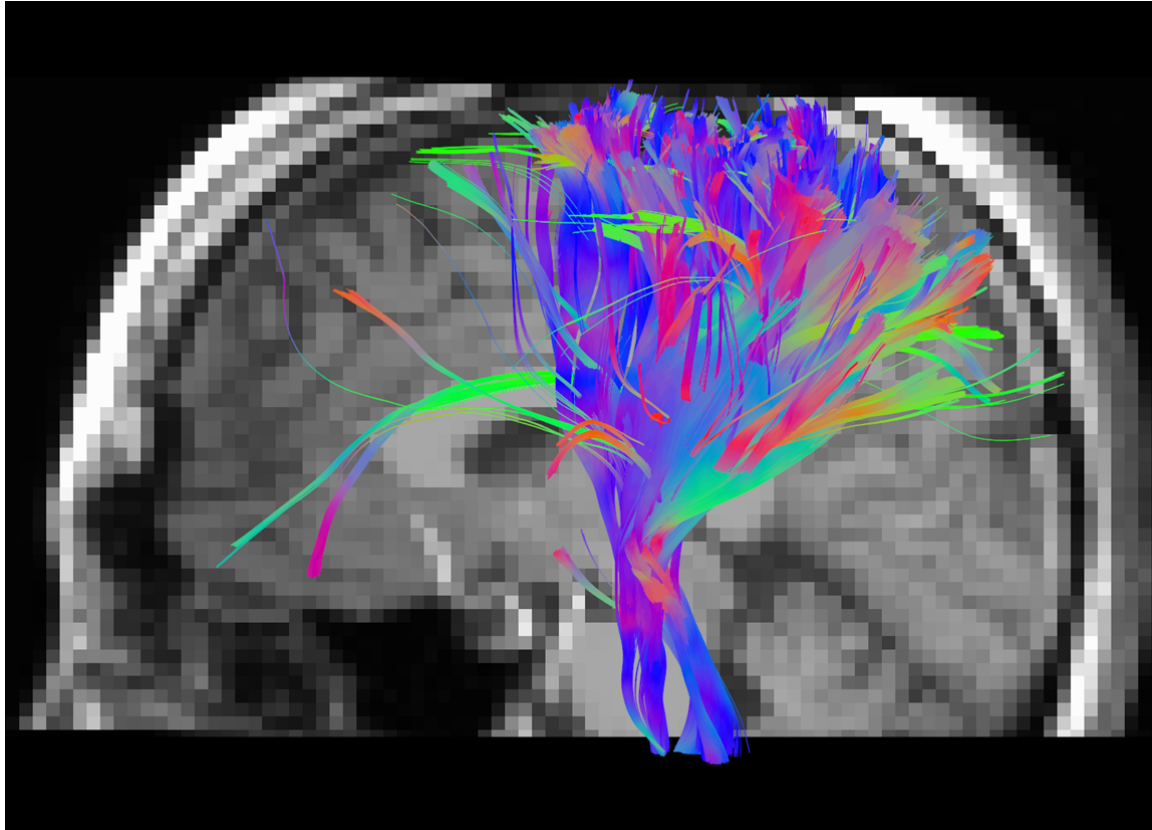


Figure 28. Top-scoring corticospinal tract results from parameter sensitivity test. Image shows the union of fibers from 4 fiber tracts tied for the highest net accuracy score of 0.93.

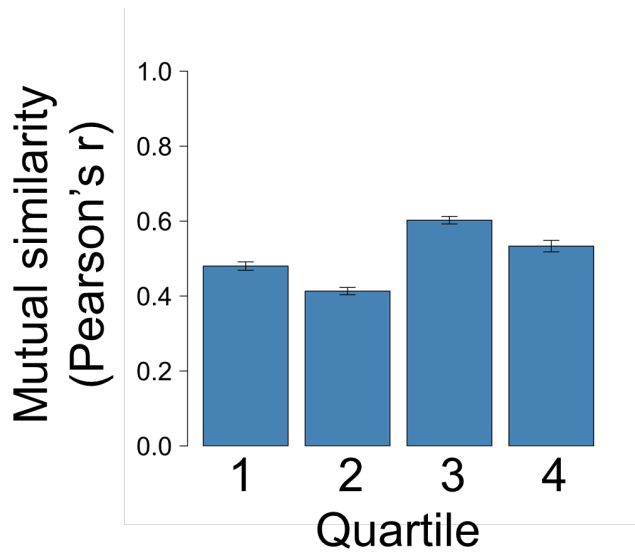


Figure 29. Mutual similarity of tractography results from Step 2 of parameter test. Bars indicate mean mutual similarity, binned by quartile in net accuracy rankings. Similarity was quantified by computing Pearson's correlations between full tract density images.

We note that of the three variables tested in the parameter sensitivity analysis, anisotropy threshold was the only one that we adjusted between datasets when tracking intercortical connections and when comparing global and targeted tracking approaches. However, the range of anisotropy thresholds observed in the high-accuracy tract group was considerably broader than the average between-session difference in tracking thresholds for a single participant, suggesting that reliability estimates in the above analyses were tolerant to small variations in tracking threshold. The mean anisotropy threshold for the top quartile of tracts in the parameter sensitivity test was 0.63, with a standard deviation of 0.19. By comparison, the average within-subject, between-session difference in the tracking thresholds used for Experiment 1 was 0.1, and the standard deviation across all datasets was 0.12.

1.3.6.3 Assessment of Spatial Error

We next addressed the contribution of spatial misalignment to between-session error. First we investigated the possibility that ROIs differed in size or shape due to between-run variability in the FreeSurfer anatomical processing pipeline (Cammoun et al., 2012). For 3 of 7 participants, we ran this pipeline twice on the same session's data, then extracted the anatomical ROIs used to construct region-to-region connection matrices. This procedure included dilation of ROIs into white matter and spatial alignment with DSI data, as described in the Methods section. We then computed the overlap of corresponding ROIs for the two runs of the FreeSurfer pipeline. Overlap for all ROIs across all three participants was 100 percent, suggesting that any randomness in the FreeSurfer pipeline was not relevant at the level of spatial precision employed in this study.

While anatomical ROIs were perfectly replicated when generated iteratively on the same dataset, they may have differed when generated from different MPAGE images for a single participant. We reasoned that if between-session differences in anatomical data (due to different patterns of T1 image distortion, for example) adversely affected reliability, then analysis in a common reference space should improve reliability. To address this question, we aligned the session 2 tracking files from the intercortical connectivity analysis to the session 1 DSI data and recomputed overlap metrics, between-session correlations of connection matrices, and detection power metrics, using the session 1 ROIs to determine connection matrices for both datasets. Following the practice adopted above, we report correlation coefficients for only the between-set intersection of non-zero connections.

Overlap metrics for the two conditions (i.e., for analysis in separate image spaces vs. aligned data) did not differ [$t(6)=0.7$, $p<0.6$]. As shown in Figure 30, correlation coefficients for both conditions were high. We compared correlation coefficients for separate vs. aligned conditions by performing a Fisher's r-to-z transformation on all coefficients (Kenny, 1987, p.274), then testing for a significant difference between these values for each participant. For 5 of 7 participants, the comparison of transformed correlation values was statistically significant at a Bonferroni-corrected alpha level of 0.05, indicating significantly better between-session correlation for aligned than separate-space data for those participants. Thus, with some individual exceptions, improvements in correlation were consistent with the possibility of between-session differences in anatomical ROI shape. Interpretation of correlation differences is further complicated by consideration of detection power metrics. The average number of connections contributing to correlation analysis in separate-space data was greater than in aligned data [$t(6)=2.6$, $p<0.05$], with a mean difference 129 connections. This difference

suggests that the set of specific connections found to be non-zero across sessions shifted, depending on whether the data were analyzed in separate spaces or the session 1 space. Given these results, spatial mismatch in the units used for reliability analysis remains a likely contributor to between-session error.

Variability in anatomical ROI definition was not relevant for the voxelwise analysis of tract density and endpoint maps. To evaluate spatial error at the single-voxel level, we computed the voxelwise correlation of aligned B0 images (Figure 31). Correlation coefficients ranged from 0.86 to 0.95, indicating an average dissimilarity (calculated as $1 - R^2$) of 0.14 (0.05), with values ranging from 0.09 to 0.25. These images were then spatially smoothed over a range of 2-20 mm FWHM, in order to compensate for partial volume effects. The results of smoothing indicated maximal similarity between the two B0 images with a smoothing kernel of 6 mm FWHM, corresponding to 1-2 voxels' worth of smoothing in every direction. At this value, dissimilarity between B0 images ranged from 0.04 to 0.21. These dissimilarity values cannot be interpreted unambiguously, as they could reflect either a failure of affine-based image alignment—that is, a failure to properly align the brain's outline across sessions—or else variability in the intensity of well-aligned voxels. However, they are consistent with the possibility that spatial error adversely affects voxelwise reliability analyses.

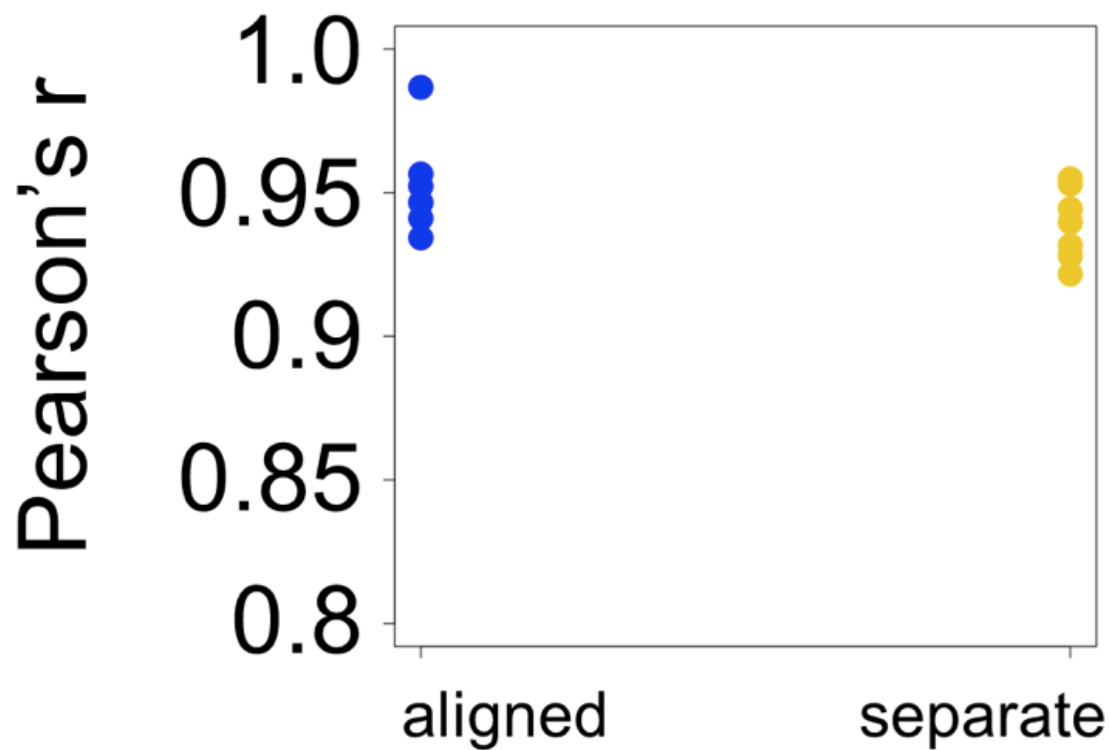


Figure 30. Reliability scores for tract pairs before and after spatial alignment. The difference in correlations suggest spatial error, possibly due to differential patterns of non-linear image distortion in T1-weighted anatomical images.

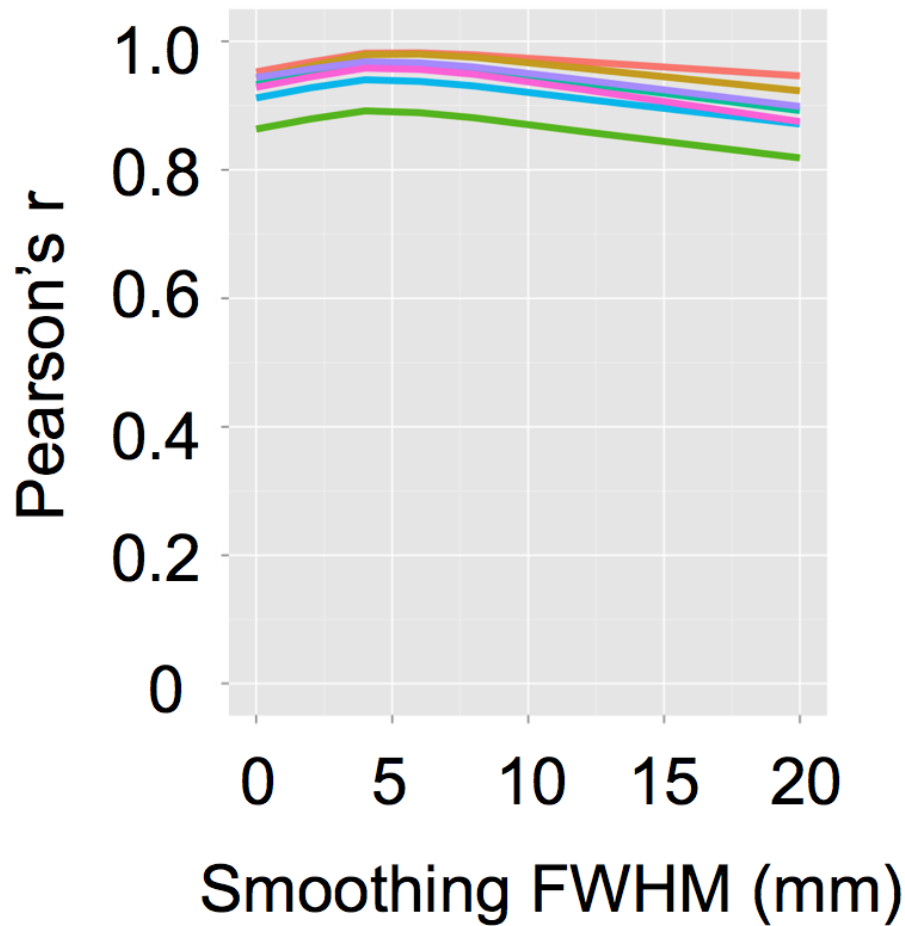


Figure 31. Quality check of affine-based spatial transformation between sessions. In order to assess residual spatial error after data alignment, we performed voxelwise correlations of the intensity values in B0 images (i.e., EPI images without diffusion weighting). We additionally investigated the effect of Gaussian spatial smoothing, to assess partial volume effects. Each line represents one participant.

1.3.6.4 Assessment of Acquisition-Related Error

The last source of error which we investigated was variability in the raw diffusion data, which could result from multiple factors, including participant head motion and hardware calibration differences. As a first-pass check of data reliability, we examined histograms of non-zero values from QA0 maps. These maps record the magnitude of the largest peak in each voxel's orientation distribution function (without regard to the orientation of that peak), and they serve as the basis for anisotropy thresholding in DSI Studio. QA0 distributions were similar in shape across all 14 datasets, skewing heavily to the left, with a single peak ranging between approximately 9500 and 16000 voxels. Paired datasets showed significant between-session differences in shape, however. Figure 32 plots the absolute value of the difference of each participant's session 2 histogram from his/her session 1 histogram. The remainders averaged 9 percent (std. dev. of 5 percent) of the total non-zero voxel count in each dataset. Differences in histogram shape were greatest at QA0 values below each dataset's mean. This rough test provides evidence of between-session variations in QA, particularly in brain areas where anisotropy tends to be lower. This characterization is highly applicable to voxels in or near gray matter and at the edge of the brain.

Next, we directly compared QA0 maps from paired sessions by spatially aligning and correlating them. As with B0 images, we applied Gaussian smoothing of 6 mm FWHM in order to compensate for partial volume effects and other small spatial errors. Table 11 compares these voxelwise correlations to B0-B0 correlations at the same smoothing level. QA0-QA0 correlations were lower than B0-B0 correlations in all participants, with an average difference of

0.137 (0.065). These results suggest the existence of between-session differences in QA0 which cannot be attributed to spatial error; rather, they are likely to result from measurement error at the time of data acquisition.

Finally, we used the spatially-aligned QA0 maps to visualize the distribution of “missing” or “masked-out voxels”: namely, voxels which survived the anisotropy threshold in one of a participant’s two DSI datasets, but not in the other. We reasoned that such cases could be at least partially responsible for the high incidence of one-off connections which we detected in the intercortical connectivity analysis. Figure 33 shows masked-out voxels for a single participant, as well as areas of greatest between-subject overlap in masked-out voxels. As predicted, these voxels were most prevalent near tissue boundaries: at the edge of the brain, in sulci and the longitudinal fissure, and around the ventricles. The spatial distribution of these voxels may support the possibility of error introduced by participant head motion.

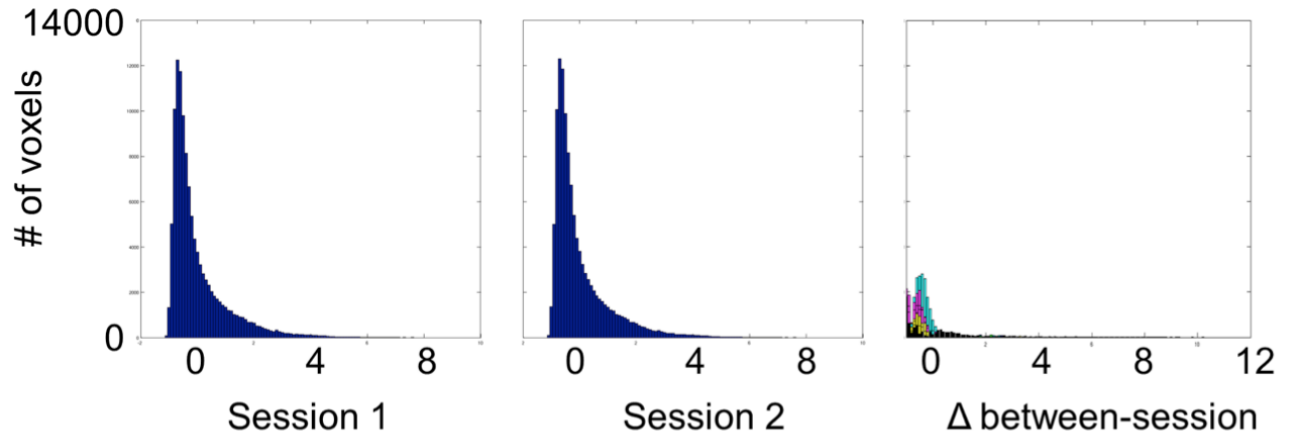


Figure 32. Between-session comparison of quantitative anisotropy (QA) maps. QA maps were the basis for thresholding tractography. Left and middle: session 1-2 QA distributions for a representative participant, plotted as z-scores. Right: Within-subject subtraction of session 1-2 QA histograms, plotted as z-scores; absolute values of the differences are plotted.

Table 11. Comparison of correlations between paired B0 images vs. paired QA0 maps.

Subject	B0-B0 correlation	QA0-QA0 correlation	Difference
0001	0.982	0.906	0.076
0004	0.980	0.871	0.109
0005	0.889	0.792	0.097
0006	0.961	0.860	0.101
0012	0.938	0.738	0.200
0013	0.967	0.845	0.122
0014	0.956	0.702	0.255
Mean (SD)	0.953 (0.032)	0.816 (0.075)	0.137 (0.065)

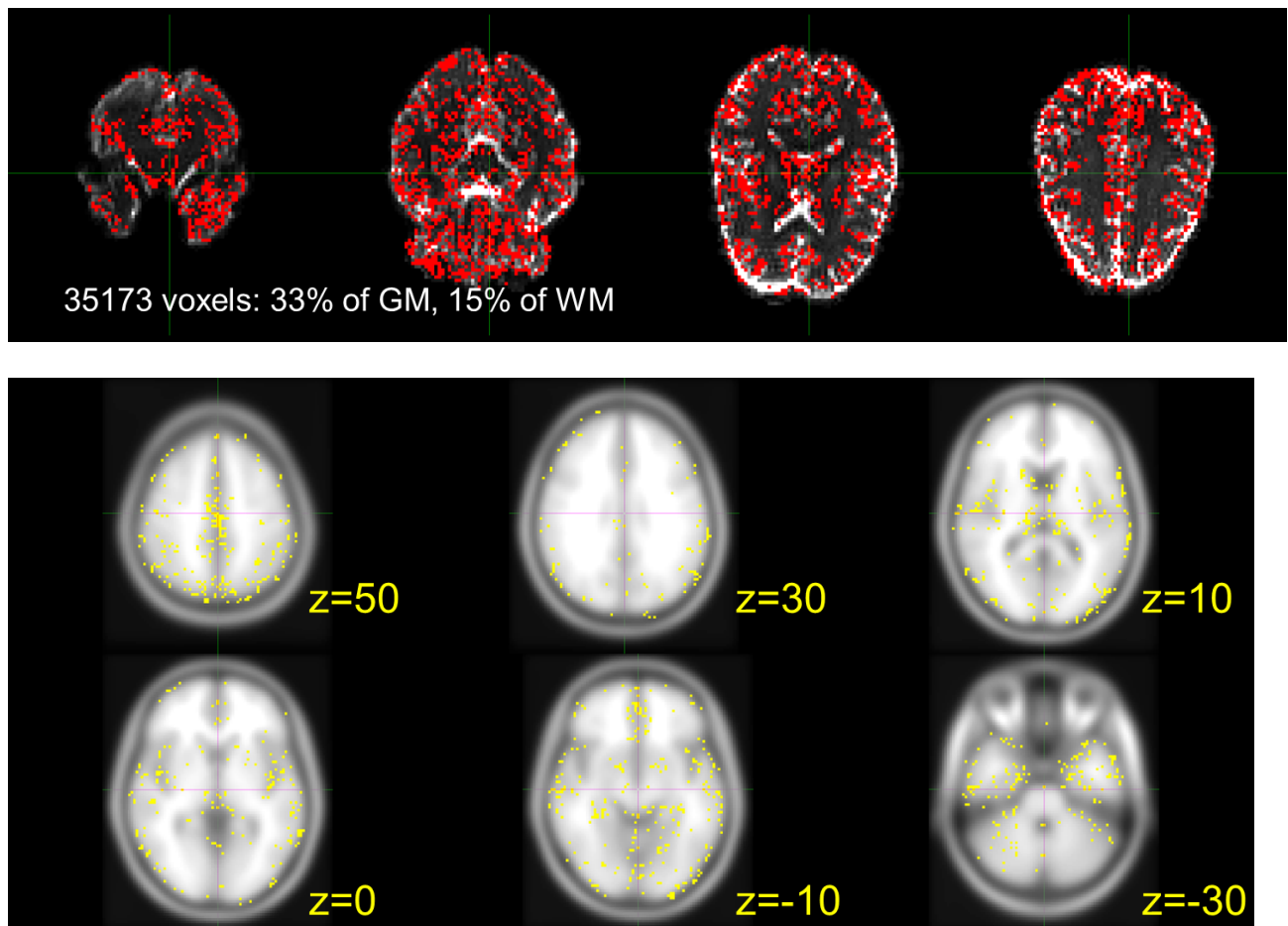


Figure 33. Voxels masked out in one session but not the other. Top: “masked-out” voxels for a single participant, representing elements of the QA0 map which survived the tracking threshold in one dataset, but not the other. Bottom: most commonly masked-out QA0 voxels, across individuals. Image shows overlap in MNI space and is thresholded to show only areas of most frequent between-subject overlap, corresponding to 3 participants.

1.4 DISCUSSION

1.4.1 Between-Session and Between-Subject Reliability

We investigated the between-session reliability of deterministic tractography based on diffusion spectrum imaging data. Reliability was investigated both on a coarse scale, by computing connection density between pairs of anatomically-defined ROIs, and on a fine scale, by voxelwise comparisons of endpoint and full-tract density maps. Importantly, we assessed the replicability of tractography results using multiple measures. Overlap metrics indicated the likelihood that a non-zero result in one session would also be present in the second, while correlation metrics indicated the reliability of fiber density values in those overlapping elements. These two metrics provided complementary information about between-session reliability.

The utility of these multiple measures is demonstrated by the intercortical connectivity analysis. Our initial correlation analyses, examining the union of non-zero connections between sessions, indicated poor reliability. (Most individual correlation coefficients were in the range of 0.3 to 0.4.) These results were at odds with the high reliability reported by Cammoun et al. (2012) using similar DSI-based tractography. However, between-session correlations in the intercortical connectivity analysis improved dramatically if we analyzed only connections present in both sessions of a participant's DSI data. The difference in correlation cannot be dismissed as a consequence of "cherry-picking" data: that is, of focusing exclusively on overlapping cells of intercortical connection matrices. Although eliminating one-off connections was expected to improve correlation coefficients, variability in connection density estimates could still have depressed reliability considerably. Instead we found very good—nearly perfect—

correlation in the fiber densities of connections which were repeatedly detected. This finding suggested that if a connection was repeatedly detected across sessions, then its density was highly reliable.

A close reading of previous reliability analyses of DSI-based tractography (Bassett et al., 2011; Cammoun et al., 2012) suggests they may have encountered a similar pattern of results. Both of these studies analyzed reliability at several spatial scales and reported high within-subject, between-session correlations of connection strength, using estimates based on fiber counts. (In the present study, we have used the method of Cammoun et al., which normalized these fiber counts by the surface area of ROI pairs and included an adjustment to address a purported length bias in deterministic tractography.) However, we note that in the current study, these high levels of correlation were only observed when restricting correlation tests to the set of connections which were repeatedly non-zero in both of a participant's DSI datasets; on average, this intersection set represented about 75% of the non-zero connections in any single dataset. Cammoun et al. appear to have found similarly noisy results: when comparing pairs of DSI datasets, the authors found a considerable number of what we have termed one-off connections (Cammoun et al., Figure 7B). Furthermore, a fraction of these one-off connections persisted even after Cammoun et al. excluded 1) connections that consisted of small numbers of fiber bundles; 2) those that were non-significant; and 3) most interhemispheric connections.

Comparison to the results of Bassett et al. (2011) is complex, but it also appears to corroborate our findings. Although the majority of Bassett et al.'s analyses deal with graph theory measures based on thresholded and unthresholded binarized matrices, they also report within-subject, between-session r^2 values well above 0.9 for DSI datasets. As in our study and that of Cammoun et al. (2012), these correlations were ultimately based on the number of fibers

connecting ROI pairs. It is unclear whether Bassett et al.'s correlations were based on thresholded connection matrices, but the apparent absence of one-off connections in the scatterplot for a single representative participant (their Figure 4B, left panel) does suggest a thresholding procedure. This interpretation is further supported by the low reproducibility of a network density measure (Bassett et al., Figure 4E). This network density measure should not be confused with connection density, as defined by this study and Cammoun et al. (2012); rather, it is a network graph property indicating the number of non-zero edges as a proportion of the number of possible edges. The low reproducibility of this measure indicates high within-subject variance in the presence or absence of particular connections across sessions; in our terminology, this result indicates relatively poor between-session overlap in binarized connection matrices.

Voxelwise analysis provided information about between-session reliability of fiber termination fields and tract bodies. We found voxelwise between-session overlap of approximately 80 percent in both analyses; these results were comparable to the overlap of connection matrices in ROI-based analysis. While voxelwise measures did not strictly reflect point-to-point connectivity, as ROI-based measures did, these results nevertheless suggest that the reliability of binarized results is relatively high. In contrast, voxelwise correlations of fiber density were much lower than in ROI-based analysis, with greater variability. The combination of these results indicates that at fine spatial scales, assessments of the presence or absence of fibers in a particular brain volume were more reliable than estimates of fiber density.

As a complement to within-subject comparisons, we examined the replicability of intercortical connections across participants. While between-session overlap for a single participant averaged about 75 percent, only 24 percent of the connections found in any single dataset were likely to be replicated across all 14 datasets analyzed. The difference in these ratios

suggests greater between-subject than within-subject variability in connectivity estimates; we note that this finding is consistent with Bassett et al.'s (2005) finding of higher intra-class correlation coefficients (ICCs) for within-subject than between-subject comparisons.

Interestingly, the likelihood of between-subject replication did not appear to depend on connection density. This result was contrary to hypothesis, as we predicted that a dense connection in a single participant would be more likely to be observed across other sessions and participants than a sparse connection. However, we found that there was only a weak linear relationship between connection density and frequency of detection, and that the distribution of one-off connections in a particular dataset spanned a similar range of connection densities as the distribution of most-often-detected (MOD) connections. This finding is methodologically important, because it calls into question the practice of thresholding brain networks on the basis of tractographic connection strength. This thresholding procedure, which has been used by Bassett et al. (2011), appears reasonable and intuitive: connections consisting of only a few diffusion streamlines are assumed to be more likely to result from random noise in diffusion data than connections that involve dense bundles of fibers. Thus, dropping the sparsest connections from a network seems to be an objective and efficient way to filter out noise, and the surviving connections are expected to be highly replicable.

However, the results of our own analyses suggest that such thresholding procedures may not work as intended (see Figures 8–9). In order to find possible explanations for this failure, it is worthwhile to consider how the behavior of both true and false connections might affect the relationship between connection density and reproducibility. First, if density actually were predictive of reproducibility, then this relationship could arise because connection density affects the rate of false positives. In this scenario, random noise in ODFs might create spurious diffusion

pathways, but the very randomness of this noise across voxels would make false connections of high density unlikely; thus, false connections would tend to be sparse. Conversely, very dense connections observed in tractography results would be likely to represent veridical diffusion pathways. Alternatively, density might affect the reproducibility of true positives. For example, sparse true connections might be more susceptible to loss or distortion of the diffusion signal, leading to a lower frequency of detection across datasets. In contrast, very dense connections would be more resistant to such noise, leading to a higher frequency of detection for dense true connections.

Hypothetical influences of connection density on hit and false alarm rates thus represent independent factors; it is possible that one or both of these factors has no predictive relationship with a connection's reproducibility. First, it is possible that false connections of high density can indeed be generated from noise in the raw diffusion data. For example, head motion by participants could create the illusion of strong diffusion in a particular orientation; importantly, this noise would be correlated across voxels, potentially leading to the detection of a false but robust diffusion pathway along the affected orientation. If realized, this scenario would contradict the assumption that spurious diffusion connections are more likely to be sparse. Our data support this possibility: in Figure 9, the distribution of connections with only 1 of 14 possible detections has a similar center and width as the distribution of MOD connections (14 of 14 detections), including similar maximum densities (right-hand tail of each distribution). These 1-time-detected connections are most likely to represent false fiber pathways, since they failed to replicate even within the paired dataset from the same individual. Second, it is possible that low connection densities do not adversely affect the detectability of true fiber pathways, and that even weak pathways can be consistently detected across individuals. Our results also admit this

possibility, since the distribution of MOD (i.e., 14-times-detected) connection densities encompasses a comparable minimum density as the distribution of 1-times-detected connections, which we have said are most likely to represent false positives. These results suggest that thresholding based on connection density may eliminate both true and false connections, and that a mixture of both connection types will survive thresholding.

A further complication in using connection density as a thresholding criterion is the ambiguous relationship between fiber density measurements and true anatomical connection strength. Current tractography algorithms make no inference about the real-world volume represented by a virtual streamline; as a consequence, no physical limitation is imposed upon the number of streamlines traveling through a given voxel or along a certain pathway. Of course, decision rules that favor stronger over weaker diffusion pathways make it likely that strong diffusion pathways will be occupied by more streamlines than weaker pathways; however, fiber density measurements will not necessarily be proportional to diffusion signal strength along white matter pathways. It is possible that our reliance on the density-based metric of Cammoun et al. (2012) obscures the relationship between true connection strength and replicability. Clarification of the precise relationship between fiber density metrics and actual diffusion volume will probably depend on empirical testing under controlled conditions, such as the use of hardware phantoms with known diffusion structure.

1.4.2 Most-Often-Detected Intercortical Connections

We found that nearly 1000 intercortical connections were detected across all DSI datasets included; as mentioned above, this number represents an average of 24 percent of the

connections detected in any single dataset. The majority of these connections were intrahemispheric, with similar proportions in the left and right hemispheres. The remaining 7.5% of connections were interhemispheric; this proportion is slightly less than (but on the same order as) Hagmann et al.'s (2008) finding of 9–14% interhemispheric fibers. The most salient brain areas in the MOD connection set included the superior and middle frontal gyri, orbital gyrus, angular gyrus, the superior parietal lobule, and lateral temporal areas (superior temporal gyrus and sulcus), among others (Tables 3–4, Figure 11). These areas were highly symmetric across hemispheres. The observed distribution of fiber lengths (Figure 15) indicates that the majority of intercortical fibers detected were of medium length (approximately 6 cm); we note that a minimum length constraint of 30 mm was imposed during tractography in order to exclude noisy connections. When we focused on brain areas involved in long-distance connections (>100 mm), we found that the brain areas of highest degree in this subset were similar to the areas of highest degree in the full set of MOD connections (Tables 3–4). Notably, however, the high-degree areas of the long-distance subset omitted such limbic areas as anterior cingulate cortex and the insula; they also included portions of occipital cortex that did not rank among the highest-degree areas in the full MOD connection network. The FreeSurfer medial wall ROI, which may have encompassed fibers terminating in subcortical structures like the thalamus, was also absent from the highest-degree areas in the long-distance subset.

Thus, we found that the most reliably detected connections—both between sessions and between subjects—involved a number of dorsal and lateral prefrontal areas, ventral and dorsal aspects of parietal cortex, and lateral temporal areas. Additionally, insular cortex, anterior cingulate cortex, and medial gray matter areas were found to have high numbers of reliable connections, primarily of short or medium distance; and occipital cortex was found to have

reliable long-distance connections. The relatively crude method used to identify these brain areas—that is, simple ranking of binarized connection counts—provides little characterization of these nodes’ place within a larger network structure. For example, it is beyond the scope of the present analysis to characterize nodes of the MOD connection network as “connector” vs. “provincial” hubs (Bassett et al., 2011; Ravasz & Barabasi, 2003); however, the high degree of these brain areas suggests that they are less likely to serve as provincial hubs, which have relatively low degree but strong connectedness within a local sub-network. In contrast, Hagmann et al. (2008) used iterative decomposition methods based on node degree and edge strength to identify core regions within an intercortical network; their results highlighted posterior midline structures including the precuneus, posterior cingulate cortex, isthmus of the cingulate, and paracentral lobule. While this set of brain areas does not overlap with our own findings, Hagmann et al. (2008) did note that superior parietal cortex, inferior parietal cortex, and lateral temporal areas—overlapping with our own results—also had high core rankings in their analysis. It is possible that further examination of our own data using graph theory techniques might resolve the apparent differences between these results. Finally, we note that our findings also overlap with Cole et al.’s (2009) analysis of highly globally-connected brain areas, based on BOLD time-series correlations. Cole et al. found high global brain connectivity in lateral and superior prefrontal cortex, rostral anterior cingulate cortex, precuneus, inferior parietal cortex, orbitofrontal cortex, anterior insula, middle occipital gyrus, and middle temporal gyrus. The present findings appear to display good overlap with Cole et al.’s (2009) findings, suggesting compatibility between structural and functional methods for brain connectivity analysis.

1.4.3 Global vs. Targeted Tracking Comparisons

We additionally compared the reliability and detection power of targeted vs. global tracking approaches. Global seeding of fiber tracking is a common technique, particularly for whole-brain connectivity analyses. By tracking with minimal ROI constraints, a researcher can efficiently generate a large number of virtual fibers throughout the brain, using a common set of tracking parameters. In contrast, targeted tracking allows a researcher to generate fibers with parameter settings optimized for detecting fibers of interest. The distinction between targeted and global tracking is not an absolute one: rather, it can be viewed as a continuum, depending upon the number and nature of ROI constraints imposed by the researcher, as well as the generality vs. specificity of parameter values chosen. In the present study, for example, corticospinal tracking involved ROIs in the brainstem, midbrain, and cortex; thus, the resulting fibers were constrained to follow a very narrow pathway up through the internal capsule, although they diverged widely above that point. Tracking could be further targeted by restricting cortical endpoints to motor, premotor, and somatosensory cortices (e.g., Verstynen et al., 2011), or it could be made more global by removing one or more ROI constraints.

The most robust result in targeted vs. global tracking comparisons was the large advantage in detection power for targeted tracking. In ROI-based analysis, targeted tracking resulted in the detection of more point-to-point connections than global tracking, for all three of the tracts investigated. In voxelwise analysis, targeted tracking was associated with the detection of many more voxels in endpoint and full-tract density maps.

Differential detection power is probably related to the smaller number of fibers, and thus poorer coverage of diffusion space, in global vs. targeted tracking files. While we generated a

uniform count of 100,000 fibers per tract in the targeted condition, fiber counts were not controlled in the global condition. As a result, AF fibers in the global condition averaged about 6.0 (3.5) percent of the total in the targeted condition; corticospinal fibers averaged 1.3 (0.5) percent; and callosal fibers averaged 9.9 (3.1) percent.

However, we feel that differential fiber counts are a further symptom of, rather than a trivial cause of, differences in detection power between targeted vs. global tracking conditions. We note that tracts in the global condition were extracted from whole-brain tracking files with a total count of 500,000 fibers. Thus, the number of seeds generated in whole-brain tracking was well in excess of the amount needed to provide detection power comparable to the targeted tracking condition. Furthermore, our detection power metrics did not depend upon specific fiber density values; rather, they only assessed the number of connections or voxels that were repeatedly non-zero across sessions. If global tracking effectively saturated the space of possible diffusion pathways, then it could be expected to detect the same connections or voxels as targeted tracking, albeit with lower fiber density.

Instead, the observed pattern of results suggests that competition among alternative diffusion pathways, coupled with the constraints imposed by the decision rules of the deterministic algorithm, resulted in a failure to detect some paths. Jarbo et al. (2012) provide a specific illustration of this phenomenon in their tracking of callosal projection fibers: global tracking failed to detect parietal and occipital projections which passed through a triple fiber crossing, but targeted tracking was successful in finding this pathway. These results suggest that even if fiber counts were increased dramatically, global tracking would still be associated with lesser detection of unique diffusion pathways than targeted tracking.

We hypothesized that targeted tracking would be associated with greater between-session reliability than global tracking. This hypothesis was corroborated by multiple findings. In ROI-based analyses, targeted tracking was associated with greater between-session overlap in binarized connection matrices; and in voxelwise analyses, between-session overlap of tract termination fields was likewise better for the targeted tracking condition. These results support the idea that tracking specific fibers of interest, with ROI constraints and tracking parameters optimized for detection of those fibers, leads to better between-session reproducibility of tractography results.

This reliability advantage is qualified by other aspects of the global vs. targeted comparisons. Comparisons of fiber density values, both in ROI-based and voxelwise analyses, suggested that correlations were statistically equivalent for global and targeted tracking conditions. This null finding indicates that targeted tracking confers no greater reliability in terms of fiber density; this result is not unexpected, given the near-ceiling correlations that we and others have detected with global tracking methods (Bassett et al., 2011; Cammoun et al., 2012). Ultimately, however, the results support the superiority of a targeted tracking approach: the detection power differences discussed above imply that targeted tracking will yield fuller termination fields and representations of region-to-region connectivity, with between-session correlations that are no worse than those based on global tracking.

We note that reliability results partially depended on the specific fiber tracts analyzed. In particular, analysis of callosal projection fibers produced a curious pattern of results. ROI-based analysis indicated significantly lower overlap of region-to-region connection matrices for callosal fibers than the AF and corticospinal tract; however, voxelwise analysis indicated the highest overlap in density maps for this tract. This apparent contradiction may have to do with

the morphology of the callosal tract and the conceptual differences between ROI-based and voxelwise analyses. The callosal fibers that we analyzed form a structure that is known to diverge widely throughout the brain, contacting most portions of the frontal, occipital, parietal, and posterior temporal lobes. (We note, however, the low detectability of fibers in the latter two areas.) Simultaneous tracking of these fibers thus necessitates a nearly global seeding approach: the most significant constraint imposed was the requirement that fibers pass through the corpus callosum to the opposite hemisphere. The lack of further constraints may be responsible for poor overlap scores in ROI-based analyses. In contrast, the sheer brain volume occupied by callosal fibers may be responsible for its high overlap scores in voxelwise analysis of density maps. Analysis of binarized callosal maps may have been less sensitive to between-session variability in region-to-region connectivity and in fiber density, relative to the AF and corticospinal tract.

1.4.4 Error Source Analyses

We performed a series of follow-up analyses to assess possible sources of between-session variability in tractography results. We considered four possible error sources: randomness in tractography seeding, variability in setting appropriate tracking parameters, spatial misalignment, and acquisition-time measurement error. Random seeding was found to have no effect on reliability when a reasonable number of seeds were generated. (This result is thought, but not known, to generalize to other tractography algorithms; we cannot dismiss the possibility that reseeding error might present a larger concern with other software packages or tracking methods.) The parameter sensitivity analysis demonstrated the robustness of tractography results to variations in tracking parameters: anatomically accurate results were observed over a wide

range of angular thresholds, smoothing/momentum values, and tracking thresholds. We note, in particular, that the observed range of tracking thresholds exceeded between-session differences in our settings of this parameter. Between-session differences in tractography are thus unlikely to relate to user settings of tracking parameters. However, the computational intensity of this analysis obligated us to limit it to a single dataset; with increases in computing power and more efficient tractography methods, we hope to repeat this analysis on a larger number of datasets and white matter structures.

In contrast to random seeding and tracking parameters, between-session errors in spatial alignment appeared to contribute to variability in tractography results. This error was evident in imperfect alignment of B0 images from each participant's DSI dataset pairs, as well as in the increased reliability of intercortical connectivity patterns for spatially aligned fiber tracts, relative to ROI-based analysis in separate image spaces. We propose that this error most likely results from differential non-linear patterns of image distortion between imaging sessions. These distortions are likely to affect automated anatomical image processing in the FreeSurfer pipeline, and thus the shape of anatomical ROIs; furthermore, they cannot be compensated for by affine-based image registration techniques. Spatial alignment errors thus affect the researcher's ability to correctly compare fiber tracts, in addition to any error they may introduce into the tractography itself.

Finally, acquisition-related measurement error appeared to be the largest factor affecting between-session reliability. This error was evident in the comparison of QA0 distributions from paired datasets, as well as in the lower correlation observed for spatially-aligned QA0 maps, relative to aligned B0 images. Variability in anisotropy measurements is unlikely to result from between-session differences in the true diffusion signal: white matter structure is slow to change,

exhibiting a gradual arc of development and degradation across the lifespan (Westlye et al., 2010). Thus, diffusion contrast is likely to remain relatively constant over time frames of weeks or months. (In this respect, DWI differs dramatically from functional imaging methods.) Furthermore, the QA metric itself does not appear to be inherently more variable than other measures of anisotropy: the between-session correlations which we report in Table 11 are comparable to prior analyses of fractional anisotropy (Pfefferbaum et al., 2003). Thus, variability in QA0 estimates appears most attributable to measurement error, which may result from a variety of causes. Post hoc, we were unable to address certain possibilities, such as between-session differences in hardware calibration. However, the spatial distribution of “masked-out” voxels—that is, voxels which survived anisotropy thresholding in one session, but not the second—was highly suggestive of head motion artifact. These masked-out voxels were most prevalent around tissue boundaries, including the surface of cortex and the ventricles.

Although it is difficult to quantitatively compare the contributions of measurement error vs. spatial alignment error, better control of participant head motion seems like the simplest and possibly most effective means of improving between-session reliability, based on these data. Currently, methods for post-hoc motion correction do not exist for diffusion MRI data collected at high b-values: insufficient anatomical detail exists in these images to allow processing with automated algorithms. However, this problem can be indirectly addressed by motion correction based on multiple B0 images, interspersed throughout the acquisition of diffusion-weighted images (Tuch, 2004). With the inclusion of such benchmark images, it may be possible to further improve data quality through post-hoc corrections, such as those typically applied to fMRI data or to DWI data collected at low b-values. However, it is unclear how translational and rotational

corrections might affect the reconstruction of ODFs in high angular resolution DWI data; empirical and model-based investigation of this question is desirable.

1.5 INTERIM CONCLUSION

Our reliability analyses yielded a mixed picture of DSI-based tractography. The fiber density for repeatedly-detected connections was highly reliable; however, we also found a high incidence of one-off connections, presumably attributable to random error in diffusion-weighted images. These one-off connections indicate significant noise in individual DSI datasets. Averaging connection matrices across sessions and/or participants appears to be a viable way to mask out noisy connections. In contrast, thresholding individual-dataset connection matrices based on fiber density does not appear to be a viable method for excluding noisy connections, as we found only a miniscule relationship between connection density and frequency of detection across datasets. Users should be aware that both averaging and density-based thresholding are likely to mask out not only false fibers, but also some valid ones, especially for connections which suffer low detectability in individual datasets. This detection problem may be particularly acute for global seeding and tracking approaches, in which the bias of decision rules in deterministic tractography algorithms causes some viable diffusion pathways to be consistently ignored. In contrast, we found that targeted tracking approaches significantly increased detection power and also provided advantages in between-session reliability, particularly when assessed in the overlap of binarized connectivity results.

Overall, the results of our study and previous investigations of DSI-based tractography (Bassett et al., 2011; Cammoun et al., 2012) urge caution in some applications of deterministic tractography. In particular, our results suggest that large-scale connectivity analyses using global seeding and tracking approaches may undersample the diffusion space of any single dataset. Furthermore, global tracking may be associated with lower reliability than targeted tracking approaches. Such global approaches have become more popular in the context of connectomic research (Hagmann et al., 2007, 2008, 2010), where they offer a computationally efficient means of estimating whole-brain connectivity. Of course, the objective of connectomic studies is ultimately to identify commonalities in brain connectivity across individuals, not to achieve maximal detection within a single dataset; thus, it can be argued that global tractography approaches represent an acceptable tradeoff between detection power and computational efficiency. However, we feel that applications of global tractography should be viewed with skepticism for now, in the absence of better quantification of hit and false alarm rates in tractography, as well as qualitative characterization of connections most likely to be incorrectly detected or rejected. Finally, we suggest alternatives to thresholding based on fiber density, such as taking the intersection of connections detected across multiple timepoints or subjects.

2.0 EXPERIMENT 2: VALIDATING DSI-BASED TRACTOGRAPHY AGAINST KNOWN ANATOMICAL CONNECTIONS OF THE FRONTAL EYE FIELDS

2.1 INTRODUCTION

Diffusion spectrum imaging (DSI) represents one technique in a broader family of diffusion-weighted imaging (DWI) methods. Coupled with fiber tractography, DSI and other high-resolution DWI techniques offer several potential advantages over invasive techniques for neuroanatomical investigation. As an MRI-based method, DWI enables the researcher to acquire comprehensive, individual-specific datasets in a single imaging session. In contrast, findings from invasive methods must be painstakingly synthesized across experiments, methodologies, and species. Moreover, because DWI is non-invasive, data about white matter structure can be acquired repeatedly on a given individual and combined with behavioral or functional imaging data.

As a relatively new technology, however, DWI requires further validation. The ideal validation method for researchers of the human brain, of course, would involve the acquisition of DWI data on living humans who have consented to postmortem donation of their brains for research purposes; invasive postmortem study could then be used to validate or challenge DWI

results. However, the ethical challenges and logistics of coordinating such research are daunting. Other possible validation methods include imaging and invasive study of non-human animals, imaging of artificial phantoms with known diffusion properties, and comparison of tractography results to prior knowledge.

In the present study, we adopted the latter approach: comparing human fiber tractography results to prior expectations of brain connectivity, based on findings from non-human primates. These findings were assessed through searches of primary and review literature, as well as by queries to the CoCoMac database (Kötter, 2004; Stephan et al., 2001). The advantage of this approach was the ability to leverage a large body of neuroanatomical knowledge gathered with so-called “gold-standard” techniques: that is, methods such as tracer injection that allow the experimenter to conclusively ascertain the presence or absence of a brain connection. We thus compared tractography results to prior findings of brain connectivity in a well-studied brain area: the frontal eye fields (FEF). This prefrontal cortical region is thought to play an important role in oculomotor control, particularly in memory-guided saccades (Lynch & Tian, 2006). In the macaque monkey, the FEF is located near the junction of the principal and arcuate sulci (Lynch & Tian, 2006). The human FEF is similarly located near the junction of the superior frontal sulcus and precentral sulcus (Nieuwenhuys et al., p. 626), and prior neuroimaging investigations have suggested that the FEF lies primarily along the anterior banks of these sulci (Rosano et al., 2002, 2003).

Of course, the process of ROI definition and hypothesis formulation is complicated by the difficulties of interspecies comparison. First, topographical differences in cortex and the localization of functionally-specialized brain areas can complicate the finding of inter-species homologies, which is necessary for using macaque findings to define human regions of interest.

Second, despite the long history of invasive neuroanatomical research in non-human primates, the space of possible brain connections in the macaque is likely to be undersampled: particular connections may be undocumented simply because they have not yet been investigated. Undersampling is a particular problem for formulating hypotheses of null connectivity. Third, and relatedly, some published null findings may be due to chance misses or to a lack of sensitivity in experimental techniques, a risk which is exacerbated by small sample sizes and infrequent replication attempts. (Of course, conservatism in invasive research is ethically desirable.) Finally, in addition to the previous three factors, real interspecies differences in brain connectivity are likely to exist. Although the frequency and nature of such differences are unknown, they could lead to detections of unexpected fiber connections, as well as failures to detect connections expected from interspecies comparison.

We attempted to mitigate these complications through procedural choices. In specifying likely target areas of human FEF projections, we limited our investigations to cortical and subcortical structures for which strong candidate homologies exist between humans and macaques. We addressed the risk posed by differences in morphology and the localization of functional areas by defining tractography masks with a wide margin of error; thus, we were unlikely to miss a hypothesized connection simply because it had not been included in the mask of possible termination points.

We were particularly cautious in formulating negative predictions about FEF connectivity, as the macaque findings that informed these could be influenced by type II statistical error. For these cases, we attempted foremost to identify macaque brain areas for which null connections to the FEF had been explicitly reported. Secondarily, we included brain areas for which no connections were apparent in tests likely to have detected them. We additionally performed

tractography at multiple levels of anisotropy thresholding, in order to confirm that observed connection densities were not a trivial consequence of tractography parameter choices.

Moreover, we refrained from making absolute predictions of null connectivity for the human homologues of these foil regions. Rather, we frame our hypotheses in terms of denser vs. sparser connectivity between high- and low-density ROI sets. We reasoned that a given connection was evolutionarily unlikely to be completely non-existent in one primate species (i.e., macaques) but densely developed in a related species (humans). We predicted that across datasets, FEF connectivity to all hypothesized high-density ROIs would be highly significant. In contrast, we predicted that connections to hypothesized low-density ROIs would be generally sparser, and that FEF connections to most of these low-density foils would not be significantly different from zero when evaluated across multiple datasets.

2.2 METHOD

2.2.1 Participant Information

Participants were the same 7 individuals (4 male, 3 female, mean age of 26.7 years) who contributed data in Experiment 1. For the current analysis, these individuals gave informed consent for additional anatomical and functional imaging at the Scientific Imaging & Brain Research Center (SIBR) of Carnegie Mellon University, Pittsburgh, PA.

2.2.2 fMRI Task Design and Procedure

Participants performed a simple block-design oculomotor task while undergoing functional imaging (Figure 34). The display consisted of a circular yellow stimulus on a black background. The vertical position of the stimulus remained constant in all task conditions, at the horizontal meridian of the display; however, the horizontal position of the stimulus changed in one of two ways, depending upon the block condition. In saccade blocks, the stimulus shifted rapidly and unpredictably from one horizontal position to another. Stimulus shifts occurred randomly, at intervals between 500 and 2000 ms. In the smooth pursuit condition, the stimulus moved smoothly back and forth between one side of the display to the other, reversing direction at an average interval of 3000 ms.

Participants were instructed to track the stimulus with overt eye movements, while maintaining stable head and neck position. Eye movements were not explicitly monitored; however, all participants had prior experience in fMRI experiments, were advised of the need to minimize head motion, and were highly compliant with other instructions. Task blocks lasted for 18 s, and each block was followed by an 18 s period of passive fixation. Data were acquired in scanner runs comprising 2 saccade and 2 smooth pursuit blocks; for each participant, we collected two scanner runs of data. The first and last halves of the run both included 1 saccade and 1 pursuit block, but the order of these was randomized across runs and individuals.

Data from the saccade/pursuit task were not collected for one participant. For this individual, functional definition of the FEF was based on data from a demanding visual attention task, the line search task. Details of this task and its implementation in the MR scanner have been previously published (Cole & Schneider, 2007). Briefly, participants must monitor multiple

spatial locations for a short line segment, oriented at a target angle. Blocks of this demanding search task, which requires high levels of attentional control, alternate with a low-load condition in which participants need only respond to a brief blink of the central fixation cross. This task has been previously shown to robustly activate the frontal eye fields (Cole & Schneider, 2007).

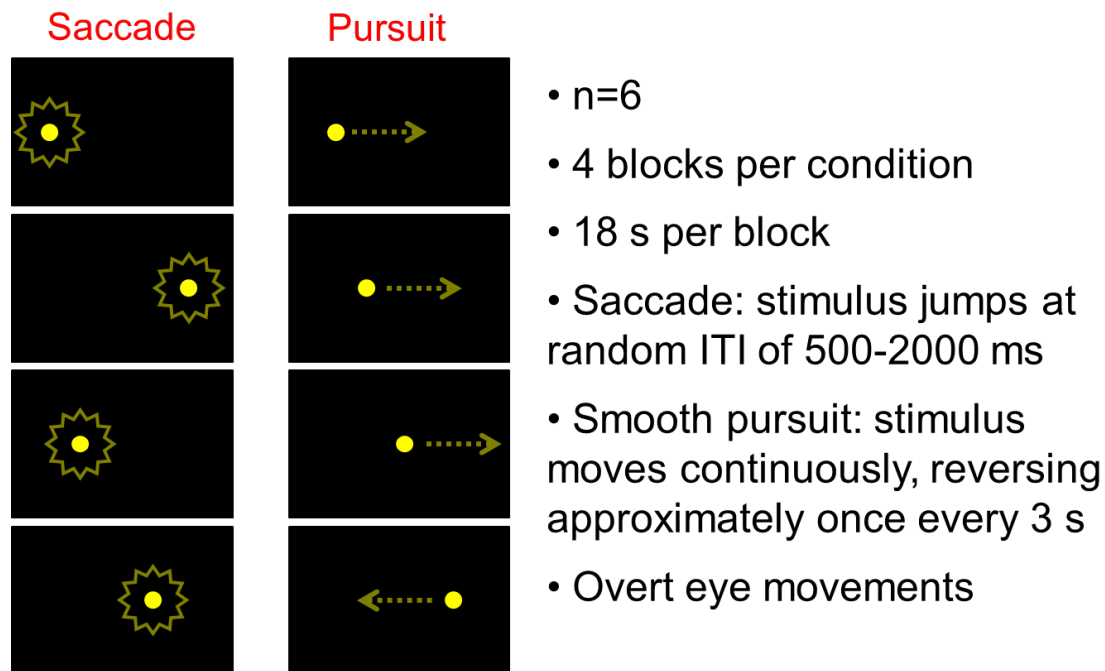


Figure 34. Oculomotor task used to localize the frontal eye fields.

2.2.3 Image Acquisition and Processing

Anatomical and functional MRI data were collected on a 3-Tesla Siemens Verio scanner with a 32-channel head coil. Anatomical imaging consisted of a sagittal T1-weighted MPRAGE

protocol with 1 x 1 x 1 mm voxels in a 256 x 256 x 176 matrix, TR = 1700 ms, TE = 2.48 ms, flip angle = 9°, and no spacing between slices. Additionally, functional MRI (fMRI) data were collected using an echo-planar imaging (EPI) sequence optimized for blood oxygenation level-dependent (BOLD) contrast. FMRI data were collected with voxel sizes of 3.5 x 3.5 x 3.5 mm in a 64 x 64 x 32 voxel matrix, TR = 2000 ms, TE = 29 ms, flip angle = 79°, and no spacing between slices. MPAGE and EPI images were aligned parallel to the plane between the anterior and posterior commissures, and slices were positioned to include the most superior aspect of cortex.

Anatomical images were subsequently reconstructed using the FreeSurfer recon-all pipeline (Desikan et al., 2006; Fischl et al., 2002, 2004), as in Experiment 1. The skull-stripped, intensity-normalized MPAGE image was used as a reference volume for alignment and analysis of fMRI data, as described below. DSI images were the same datasets employed in Experiment 1: each of the 7 participants was represented by 2 DSI datasets. As in reliability analyses, DSI data were reconstructed using the generalized q-sampling imaging method (GQI; Yeh et al., 2010).

FMRI data were pre-processed using the AFNI/SUMA (Cox, 1996; Cox et al., 1997) and FSL (Smith et al., 2004; Woolrich et al., 2009) software packages. BOLD timeseries were first corrected for slice-by-slice differences in time of acquisition. Subsequently, head movement parameters were calculated using a mean EPI volume as the reference image. We next calculated a 12-point affine transformation between the EPI reference and the processed MPAGE image output by FreeSurfer. This transformation was concatenated with the motion correction parameters for each EPI volume and simultaneously applied, in order to prevent unnecessary data blurring from multiple spatial interpolations. The processed, spatially-aligned EPI time-series were then projected onto the cortical surface; the value at each surface node represented an

average over 10 steps along the segment projecting normal to the cortical surface over the thickness of the cortical gray matter. The time-series for each surface node was normalized on a run-by-run basis to a mean value of 100. Finally, functional data were smoothed on the cortical surface to a target full-width half-maximum (FWHM) of 4 mm.

2.2.4 Deterministic Tractography Methods

As in Experiment 1, we performed deterministic tractography in DSI Studio, using Euler fiber propagation, sub-voxel random seeding, and random choice of initial fiber orientation. We employed the same angular threshold and smoothing values that were used in targeted tractography (70° and 0.8, respectively). The minimum length constraint was set to 20 mm, in order to ensure detection of short-range connections between the FEF and supplementary eye field (SEF), located proximal to one another. The maximum length constraint was set to 120 mm, in order to allow detection of the most distal FEF connections, but to exclude false fibers that might follow improbable looping trajectories throughout the brain.

We thresholded participants' quantitative anisotropy (QA) maps using the same tracking thresholds used in Experiment 1. However, we calculated additional low and high thresholds around this starting level, so that we could examine the consistency of tractography results over a range of threshold values. Low and high thresholds were determined by subtracting or adding one half of the standard deviation of non-zero QA values within each dataset (Table 12).

Seed and target ROI masks were created using a process similar to that of Experiment 1. As above, cortical ROI masks were created by dilating surface representations of each ROI below the white matter surface by 1 x cortical thickness, saving them as binarized MRI volume images,

and spatially aligning them with DSI data. Spatial alignment was based on the previously-computed affine transformation matrices between high-resolution MPRAGE images and DSI B0 images; alignment was performed using FSL's FLIRT program with trilinear interpolation and no spatial blurring. After alignment, a clipping threshold of 0.1 was applied to each ROI mask to prevent the inclusion of voxels that were only minimally occupied by the ROI, thus minimizing partial volume effects. Subcortical ROI masks were extracted from the FreeSurfer aparc+aseg volume and spatially aligned using the same process applied to cortical ROIs.

Table 12. Low, medium, and high QA thresholds for FEF tractography. Medium thresholds are the same as those used in Experiment 1. Low and high thresholds were obtained by subtracting or adding half a standard deviation to medium thresholds.

Participant	Session	Low	Medium	High
1	1	0.1634	0.5000	0.8366
1	2	0.2350	0.5761	0.9172
2	1	0.1878	0.6274	1.0670
2	2	0.0998	0.4692	0.8386
3	1	0.1265	0.3631	0.5997
3	2	0.0932	0.3101	0.5270
4	1	0.1409	0.3756	0.6103
4	2	0.0634	0.3086	0.5538
5	1	0.1858	0.5449	0.9040
5	2	0.3136	0.7201	1.1266
6	1	0.1818	0.4585	0.7352
6	2	0.1483	0.3792	0.6101
7	1	0.3215	0.5917	0.8619
7	2	0.2586	0.5195	0.7804

2.2.5 Macaque Meta-Analysis and Human ROI Definition

We defined the FEF through statistical contrasts in fMRI data, based on the saccade/pursuit task (6 participants) and the line search task (1 participant). Masks of the left and right FEF were used as tractography seed points; these masks are described in greater detail in the Results section.

We next reviewed published findings of macaque brain connectivity, with the goal of identifying possible high-density target and low-density foil areas for tractography of FEF fiber projections. We specifically hoped to identify regions of interest for which approximate human homologies could be identified, and which could be defined on the basis of anatomical landmarks. FEF projections in the macaque were first identified from review papers and from primary literature. For example, the review of Lynch & Tian (2006) identified the pontine nuclei, thalamus, superior colliculus, caudate nucleus, and putamen as the subcortical structures most strongly connected to the FEF. Notable cortical inputs included functionally-related oculomotor control areas: namely, the parietal, prefrontal, and supplementary eye fields, as well as the medial superior temporal area and area 7m (Lynch & Tian, 2006). Lynch & Tian note that parietal connections form the most prominent of these reciprocal links with the other cortical eye fields; additionally, they highlight Barbas & Mesulam's (1981) report of FEF projections to areas V2, V3, and V4, as well as Brodmann's areas 18 and 19. FEF connections to occipital and temporal visual association areas are corroborated by Barbas et al. (1988), who found heavy projections to the rostral calcarine fissure and caudal superior temporal sulcus, as well as the occipito-temporal, parieto-occipital, inferior occipital, and lunate sulci. Barbas and colleagues also corroborated the presence of heavy projections to lateral portions of the intraparietal sulcus.

We attempted to further verify these connections through a query to the CoCoMac database, which integrates findings about macaque brain connectivity over a diversity of experimental methods and anatomical classification schemes (Stephan et al., 2001; Kötter, 2004). Importantly, findings recorded in the database rate each entry for the precision of the original report and the density of the reported connection, when specified or apparent. Connection densities are rated with a zero, a number from one to five, or with an “X”. In this scheme, zero density indicates an explicit report of null connectivity; values from 1 to 5 indicate a coarse scale from sparsest to densest connectivity; and an “X” indicates that a given connection was reported or apparent, but that its density could not be determined.

A connectivity search using the keyword “FEF” yielded 169 results of specified density and 91 results of unspecified density. The FEF was identified in these results either eponymously, or as area 8a in the scheme of Walker & Walker (1940). Query results (Table 13) confirmed the presence of FEF connections with the thalamus (findings from Goldman-Rakic et al., 1985; Olszewski, 1952; Romanski et al., 1997; Stanton et al., 1989; and others), caudate nucleus (Shook et al., 1991), and putamen (Shook et al., 1991). Cortical projection areas with quantified connections included the mid-anterior and mid-posterior cingulate cortices (BA 23, 24), inferior and superior temporal cortices, the SEF, premotor cortex, and a wide range of parietal areas. Additional connections of unspecified density connected the FEF to dorsolateral prefrontal cortex (BA 8, 9, and 46) and to visual association areas in extrastriate and parieto-occipital cortex. No findings of collicular or pontine connections were returned by our query; however, we treated these connections as verified, on the basis of Lynch & Tian’s (2006) review. That survey noted that both superior colliculus and pontine connections were documented by Huerta et al. (1986) using a combination of microstimulation and tracer injections; additionally, it reported

independent confirmation of FEF-pontine connections by Stanton et al. (1988), Leichnetz et al. (1984), and others.

Based on these findings, we chose subcortical targets for FEF tractography in the striatum (caudate nucleus and putamen), thalamus, and superior colliculus (Figure 36). Most high-density target ROIs were based on volumes created by the FreeSurfer automatic parcellation and segmentation scheme; however, the superior colliculus was hand-traced for each dataset on an MPAGE anatomical image co-registered to the DSI data. We included mid-anterior cingulate cortex as an approximate homology to the macaque cingulate motor area (CMA), and the posterior superior frontal gyrus in order to represent the SEF. Because of the wide spatial distribution of parietal connections, and to compensate for ambiguities in finding macaque and human homologies, we deliberately included a large volume of parietal cortex in our tractography mask, comprising the angular and supramarginal gyri, intraparietal sulcus, superior parietal lobule, precuneus, and subparietal sulcus.

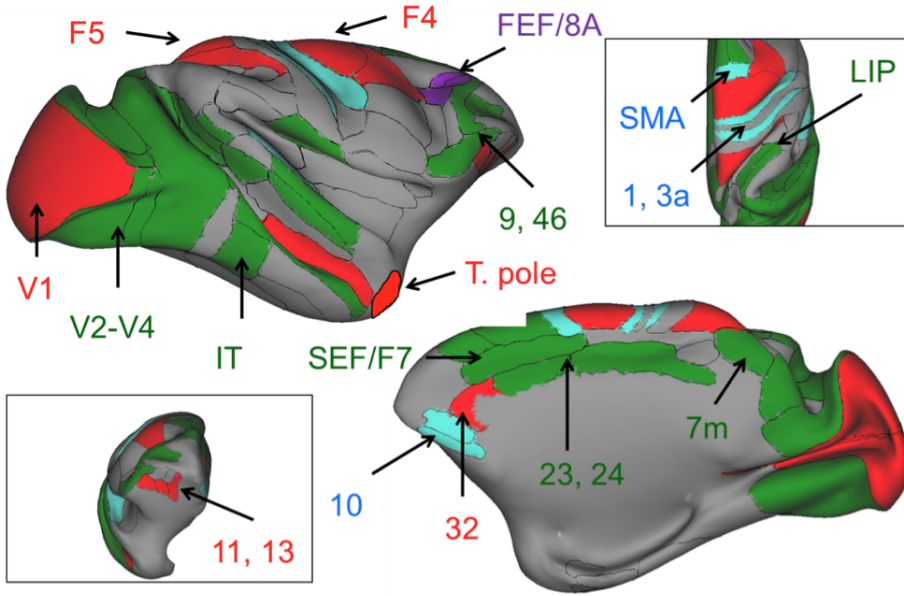


Figure 35. Map of target and foil FEF cortical projection areas. FEF connectivity was assessed through CoCoMac-based meta-analysis. FEF is labeled in purple; target regions are labeled in green; and foil regions are labeled in red. Cyan-colored regions indicate ambiguous or conflicting evidence for FEF connectivity. All area names refer to the macaque brain atlas of Lewis & Van Essen (2000). SMA: supplementary motor area; LIP: lateral intraparietal area; IT: inferotemporal cortex. Subcortical foils and targets are not shown.

Table 13. Summary of CoCoMac query of FEF connectivity. Connection density is either unspecified, or else rated from 0 to 5: a report of 0 indicates a reported null connection, while a report of 5 indicates very dense connectivity. Entries in red indicate explicitly null connections. Entries in blue indicate conflict between explicit reports of null and positive connections.

Region	Number of observations	Mean density (0-5)	Std. dev.	Add'l. observations, unspecified density
Amygdala	9	0	0	
Auditory cortex	18	0.17	0.38	4
Caudate	7	3	0	
Clastrum	2	1.5	0.71	
Inferior temporal	3	1.33	1.15	6
Insula	1	0	0	1
Limbic cortex				
23	5	0.8	0.45	1
24	5	0.8	0.45	
25	1	0	0	
32	2	0	0	
CMA	2	0.5	0.71	2
Lateral PFC				
8				2
9				1
46				7
Middle/superior temporal	13	0.92	1.44	13
Orbitofrontal cortex				
10	2	0	0	1
11	1	0	0	
12				1
13	1	0	0	
14	1	0	0	
orbital	1	0	0	
proisocortex				
Parietal cortex	11	0.91	1.30	9
Parieto-occipital				6
Peristriate/extrastriate				13
Premotor	7	0.57	0.79	5
Primary motor	12	0	0	1
Primary visual	3	0	0	

Table 13 (continued)				
Putamen	9	3	0	
SEF	2	3	0	
Somatosensory				2
Substantia nigra	1	1	0	
Temporal pole	1	0	0	
Thalamus	37	1.46	1.04	11

We did not include lateral prefrontal cortex as a region of interest, due to the difficulty of forming hypotheses about FEF connection density in this area of the brain. The prefrontal cortex may represent the region of greatest elaboration in human beings, relative to non-human primates (Hill et al., 2010; Orban et al., 2004; van Essen, 2004); this expansion complicates interspecies comparison. This problem is exacerbated by the fact that connections between oculomotor control areas in lateral PFC and the FEF have received relatively little investigation: we note that in CoCoMac query results, none of the reported lateral PFC connections included assessments of connection density. In contrast, a number of quantified and unquantified connections were reported to various parts of parietal cortex. We also chose not to test visual association and oculomotor control areas in extrastriate cortex, inferior temporal cortex, and superior temporal cortex, because these targets could not be well defined using anatomical ROIs in the FreeSurfer segmentation/parcellation scheme.

Specification of low-density foil ROIs for FEF tractography was more complicated, as absence of evidence could not be taken as evidence for null connectivity. To the extent possible, we attempted to define ROIs on the basis of explicit negative reports of FEF connectivity. Defining foils was also difficult due to the relative coarseness of FreeSurfer-defined anatomical ROIs. As discussed above, low-density foil areas were thus not intended to represent absolute

hypotheses of null connectivity; however, we predicted that they would have sparser connections with the FEF than did target ROIs.

On the basis of the CoCoMac query results (Table 13), we included the amygdala as a foil region, given multiple reports of null connections between its constituent nuclei and the FEF. We also included the globus pallidus as a foil region, as we could find no evidence of reported connections with the FEF either in CoCoMac query results or in a broader literature search. Although FEF connections to the caudate nucleus and putamen are well established, review papers that discussed them made no mention of pallidal connections (Harting & Updyke, 2006; Lynch & Tian, 2006). We note that this pattern of connectivity is consistent with a general theory of computational loops connecting cortex, the basal ganglia, and the thalamus (Narayanan, 2003), in which the globus pallidus is only indirectly connected to cortex and tonically inhibits thalamic input to cortex. We were similarly unable to find any reports of FEF connectivity to the hippocampus; thus, this structure was also included in the foil set.

Cortically, we included the anterior cingulate cortex as a low-density foil, representing an approximate homologue of Brodmann's areas (BA) 25 and 32. Additionally, we included frontomarginal and frontopolar areas, given explicit reports of null connectivity to BA 10. We note, however, that the evidence for null connectivity in BA 10 is apparently contradicted by one report of an unquantified connection to this area, possibly indicating differences in methodology or taxonomy across studies. Similarly conflicting evidence existed for the insula; however, we chose to include this region in the foil set, reasoning that weak support for its connectivity to the FEF implied, at most, a sparse connection. Given several reports of null connectivity to primary motor areas, and the sparsity of evidence for projections to somatosensory cortex, we also included the central sulcus and postcentral gyrus as foil regions. We note, however, that

CoCoMac results included two reports of unquantified connections between BA 8 (approximate location of the macaque FEF) and somatosensory cortex, to BA 1 and 3a (Barbas et al., 1988). In light of multiple reports of null connectivity to macaque area V1, we included the occipital pole and calcarine sulcus as low-density foils; however, we note that overlap of these anatomically-defined ROIs with functional areas V2, V3, or V4 could result in some detection of FEF fiber projections, most likely near the borders of those ROIs. We also included portions of the transverse temporal gyrus, transverse temporal sulcus, and mid-anterior superior temporal gyrus in the foil set, due to multiple reports of null FEF connectivity to auditory cortex. However, we note that a minority of observations (3 of 18, all from Hackett et al., 1998) did report positive connections in the vicinity of the caudal auditory belt; given these findings, we considered sparse connectivity to this part of temporal cortex as a possibility. Finally, we included the temporal poles as foils, given a previous report of null connectivity.

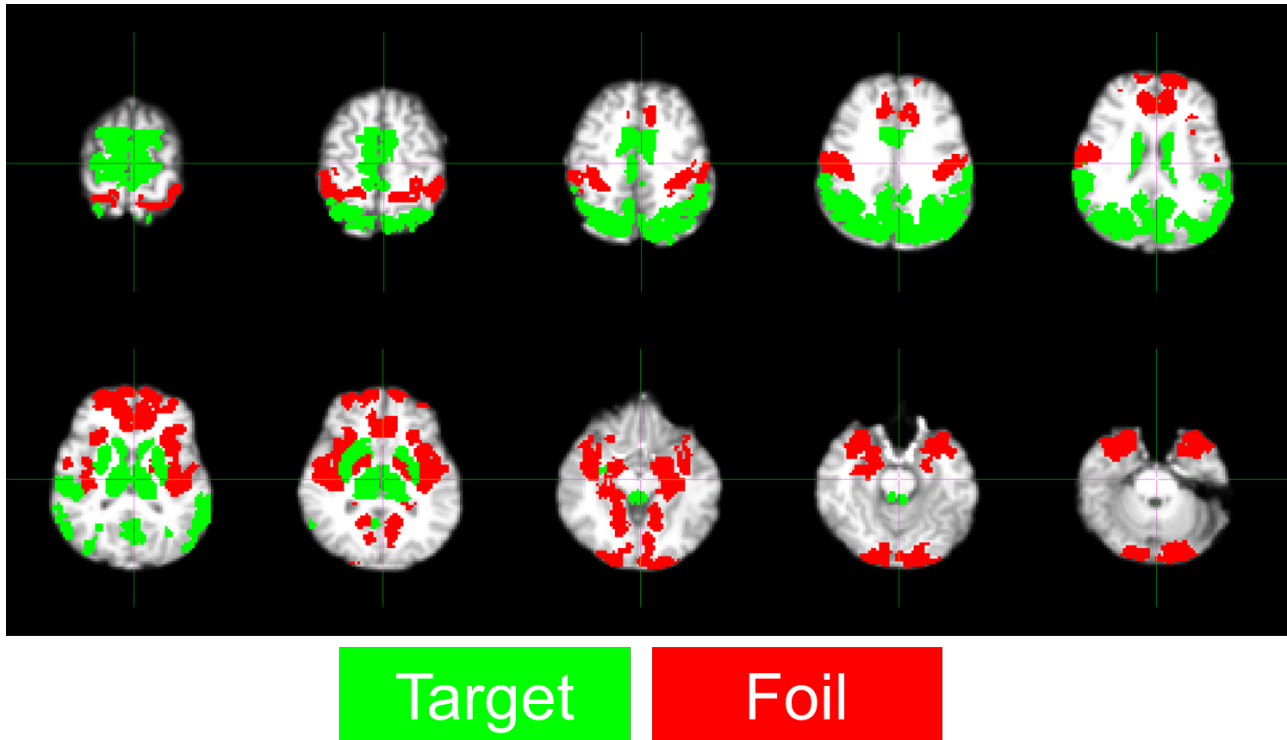


Figure 36. High-density target and low-density foil masks for FEF tractography. Target and foil ROIs are green and red, respectively. Masks are shown for one representative participant. FEF ROIs are not shown.

2.2.6 Statistical Methods & Hypotheses

We compared the fiber projections to ROIs in the high-density target and low-density foil set using a combination of repeated measures ANOVAs and post-hoc one- and two-sample t-tests. Omnibus ANOVAs were performed separately for each level of tracking threshold tested (low, middle, and high); they included fixed factors of ROI group (high-density targets vs. low-density foils) and hemisphere (left/right), plus a random factor of participant. We predicted greater fiber density for targets than foils. No effect of hemisphere or interactive effect was predicted. We

performed subsequent one-sample t-tests on fiber density values for each ROI tested versus a null hypothesis of zero density, collapsing across hemisphere and participant. Due to the large number of tests involved (30 ROIs), a Bonferroni correction was applied to adjust the familywise error rate. We predicted that connection density for all targets would be significantly greater than zero; in contrast, we predicted that connection density for many foil ROIs would be less reliable. Finally, we conducted post-hoc tests to compare the density of each connection from the target set to the total distribution of fiber densities in the foil set; we predicted that individual target connections would be significantly greater than foil connections. As above, a familywise error rate of $\alpha=0.05/30$ was employed to compensate for multiple comparisons.

2.3 RESULTS

2.3.1 ROI Definition

Tractography seed masks were created by defining the FEF through statistical contrasts in fMRI data. For 5 of 7 participants, the left and right FEF were defined through a contrast of all task conditions against a resting baseline in the saccade/pursuit localizer task. This contrast was thresholded at a significance level of at least $p<0.01$, corrected for multiple comparisons using the false discovery rate technique (Benjamini & Hochberg, 1995; Logan & Rowe, 2004); an exception to this criterion was made for defining the right FEF in one participant, and an FDR-corrected level of $p<0.05$ was used instead. For the sixth participant in the saccade/pursuit task,

the contrast of all blocks vs. baseline produced noisy results, unsuitable for functional localization of the FEF; instead, the FEF was defined from the overall F statistic for the general linear model, thresholded at an FDR-corrected significance level of $p < 0.001$. For the remaining participant, the FEF was defined from the contrast of high-control search vs. the low control baseline in the line search task (Cole & Schneider, 2007), thresholded at a corrected value of $p < 0.001$.

Previous functional imaging investigations of the human FEF have localized it to the precentral sulcus (particularly the anterior bank) and the posterior portion of the superior frontal sulcus (Rosano et al., 2002, 2003). In all datasets, the FEF was identified as the cluster of functional activation at this location. In some cases, FEF activation was continuous with motor-related activation in the precentral gyrus and with activation of the supplementary eye field (SEF), located on the medial wall and most dorsal aspect of the posterior portion of the superior frontal gyrus. In these cases, anatomical masks of the SFG and precentral gyrus were used to mask activation maps and isolate the core of the FEF.

Table 14 reports volumes of the final left and right FEF ROIs for all datasets, as well as the ratio of left vs. right FEF volume. As expected, the between-session, within-subject correlations of FEF volume were high [$r = 0.89$, $p < 0.0001$, 95% confidence interval = 0.74, 1.0]. This correlation was unsurprising, since FEF masks for paired DSI datasets were based on the same fMRI data, collected in a third imaging session. Less expected were the hemispheric differences in ROI volume: in all but one dataset, right FEF activation was more extensive than in left FEF, creating a reliable hemispheric bias [$t(13) = 4.8$, $p < 0.001$]. Although unanticipated, this effect is consistent with a theorized right-hemisphere bias in attentional and perceptual processing.

Table 14. Mask sizes for left and right FEF.

Participant	Session	Left FEF voxel count	Right FEF voxel count	Left/right ratio
1	1	539	998	0.54
1	2	563	962	0.59
2	1	257	443	0.58
2	2	241	210	1.15
3	1	480	494	0.97
3	2	576	630	0.91
4	1	383	489	0.78
4	2	391	485	0.81
5	1	333	762	0.44
5	2	323	701	0.46
6	1	368	438	0.84
6	2	372	461	0.81
7	1	432	530	0.82
7	2	294	471	0.62

We additionally compared the volumes of ROI masks for high- and low-density sets, since a large discrepancy would raise the possibility that overall differences in fiber density were simply due to the difference in mask size. Figure 37 shows the mean across sessions and hemispheres of total target mask vs. total foil mask volume, with standard error bars. Target masks were reliably larger than foil masks [$t(27)=4.9$, $p<0.0001$]. However, this effect was relatively modest: the mean difference of 803 voxels was equivalent to approximately 10% of total foil mask volume. Still, in later analysis we considered whether the difference in mask volumes could have biased the comparison of fiber densities for target vs. foil ROI sets.

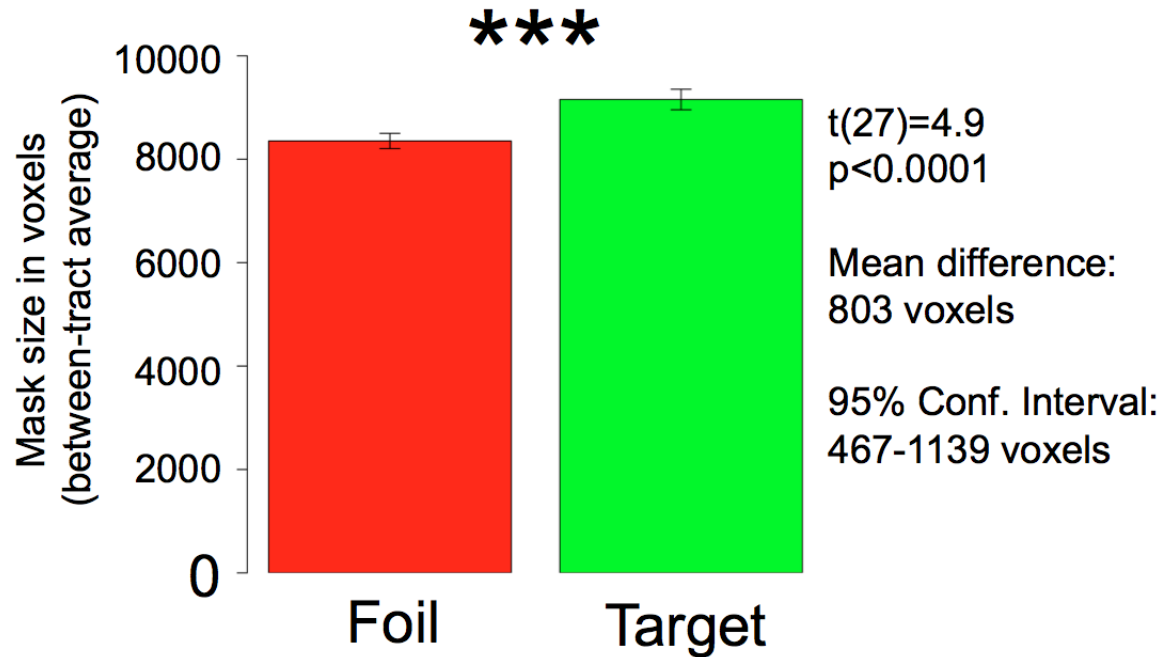


Figure 37. Combined mask sizes for high-density targets and low-density foils.

2.3.2 Target vs. Foil Comparison

We first tested the overall density of target vs. foil ROI connections at the medium anisotropy threshold, corresponding to the thresholds used in Experiment 1. An ANOVA with fixed factors of ROI group and hemisphere, as well as a random factor of participant, was conducted on fiber density values. This analysis indicated a main effect of ROI group [$F(1, 830)=42.5$, $p<0.0001$]. Neither the main effect of hemisphere [$F(1,830)=0.1$, $p<0.8$] nor the interaction of ROI group with hemisphere [$F(1,830)=2.5$, $p<0.2$] were significant. As previously noted, the total volume of ROI masks for the high-density target set was consistently larger than in the low-density foil set, raising the concern that tractography could have been biased to detect

in the low-density foil set, raising the concern that tractography could have been biased to detect fibers projecting to target ROIs. However, the observed pattern of results makes this artifactual explanation unlikely. The mean density of projections to target ROIs was 26.5 (41.5), compared to 5.0 (6.2) for foil ROIs; this difference is disproportionately larger than the difference in tractography mask volumes.

We also found a consistent hemispheric asymmetry in the volume of FEF ROI masks, as described above. We thus asked whether or not an effect of hemisphere could be found upon simple fiber counts, unnormalized by ROI volume. However, an ANOVA on fiber count values produced the same pattern of effects as fiber density analysis: although the effect of ROI group was highly significant [$F(1,830)=101.2$, $p<0.0001$], we again found null effects of hemisphere [$F(1,830)=0.0$, $p<0.9$] and condition by hemisphere [$F(1,830)=0.6$, $p<0.5$].

Post-hoc tests on the low-density foil set revealed only two connections that were significantly different from zero, after correction for multiple comparisons: projections from FEF to the central sulcus and to the postcentral gyrus. These projections most likely indicate connections with primary motor and somatosensory cortices. Although connections to homologous areas in the macaque have not been widely reported, they have precedence in Barbas et al.'s (1988) finding of connections from macaque area 8 to areas 1, 3a, and 4c. Connection density values for other ROIs exhibited significant variability across datasets: notably, the insula, globus pallidus, and hippocampus received dense FEF projections in some datasets, but no projections in others. This variability is discussed below in greater detail.

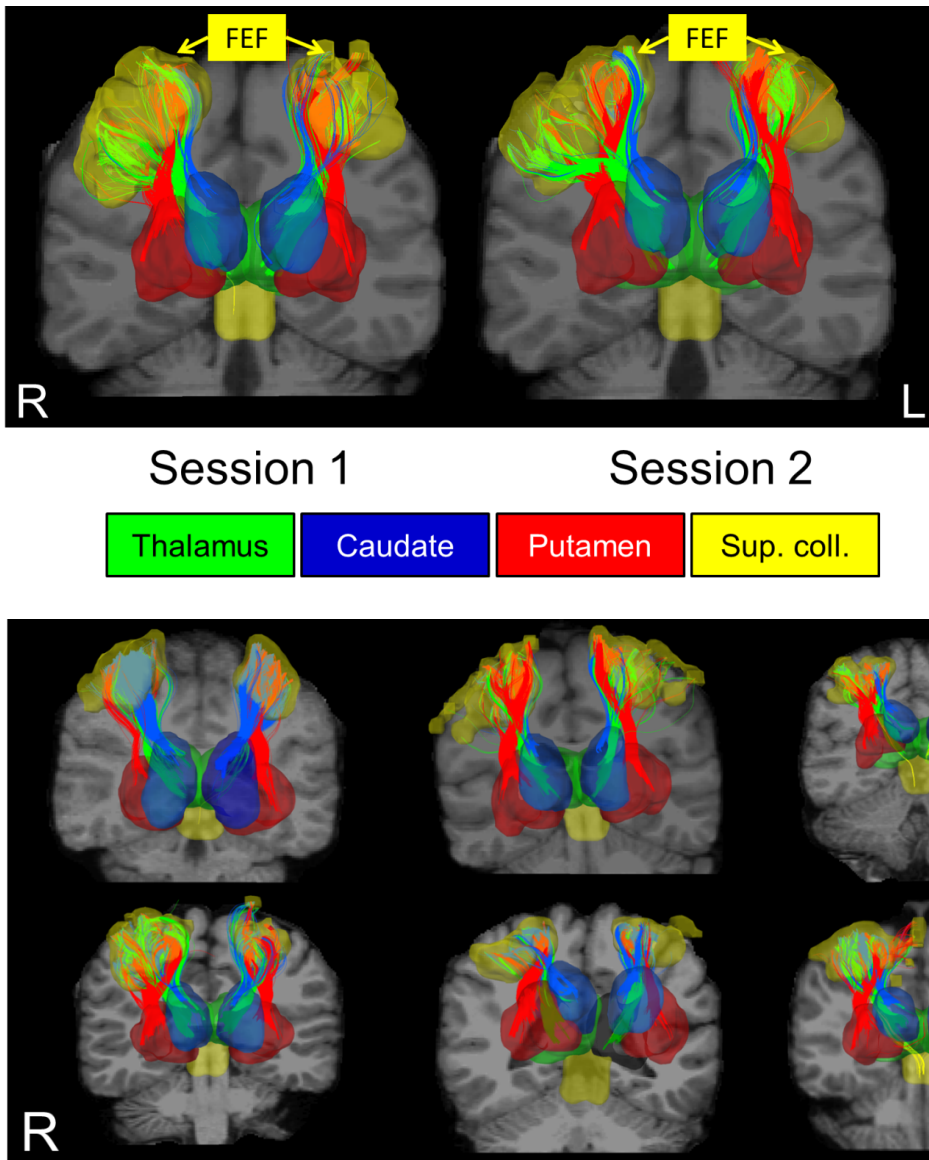


Figure 38. Fiber connections from FEF to subcortical targets. FEF is shown in yellow, thalamus in green, putamen in red, and caudate in blue) Top: coronal view of representative tracts for participant 1. Session 1 results are on the left; session 2 results are on the right. Images are presented in radiological orientation. Bottom: coronal view of subcortical projections for participants 2–7. For these 6 participants, results are shown for only the first session; however, tractography and subsequent analysis were performed in all 14 available DSI datasets.

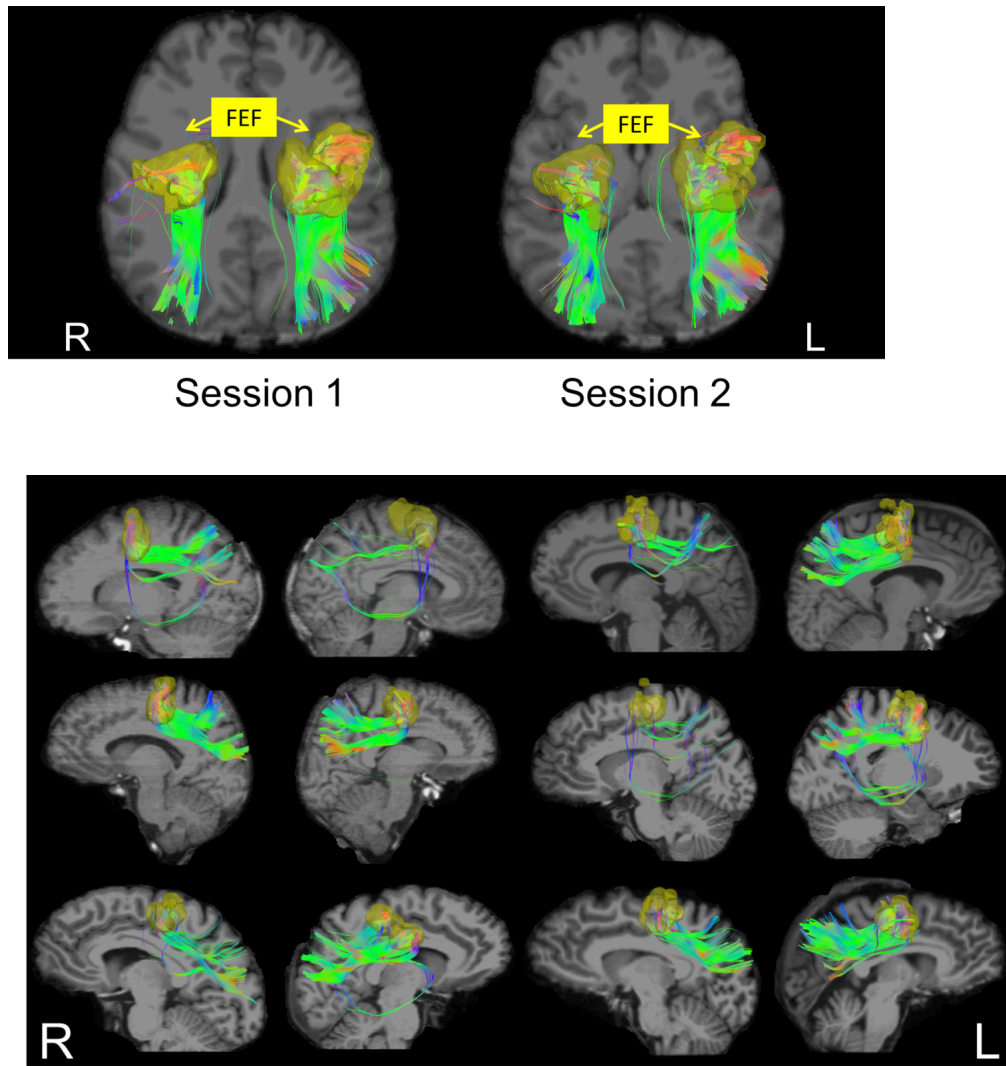


Figure 39. Fiber connections from FEF to posterior parietal cortex. Top: axial view of representative tracts for participant 1. Session 1 results are on the left; session 2 results are on the right. Images are presented in radiological orientation, with the left hemisphere displayed on the right side of each image. Bottom: sagittal view of PPC projections for participants 2–7. For these 6 participants, results are shown for only the first session; however, tractography and subsequent analysis were performed in all 14 available DSI datasets.

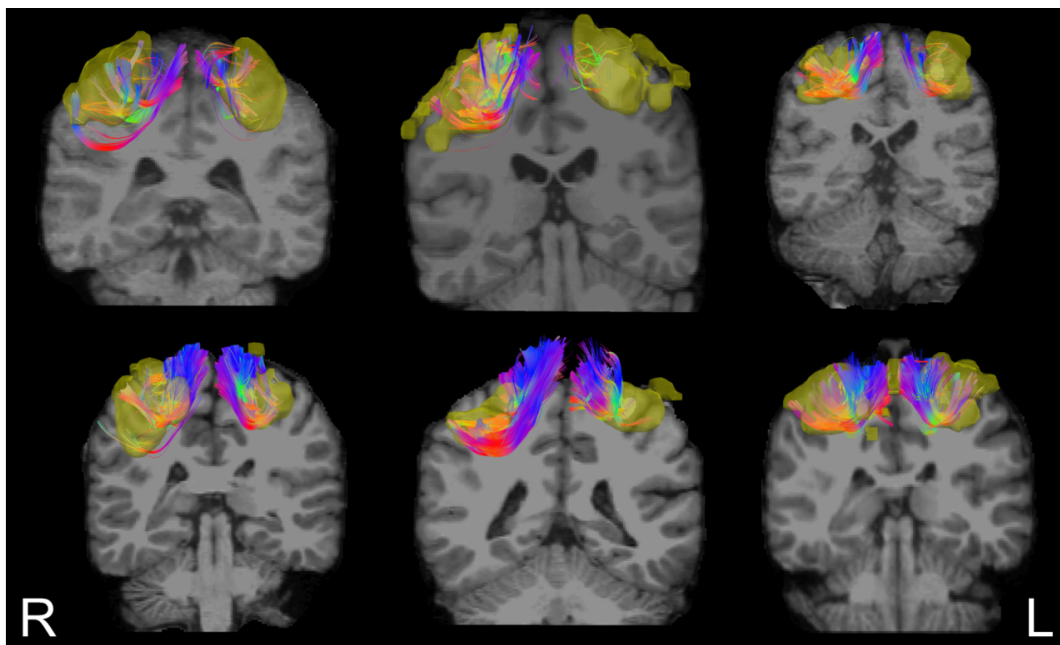
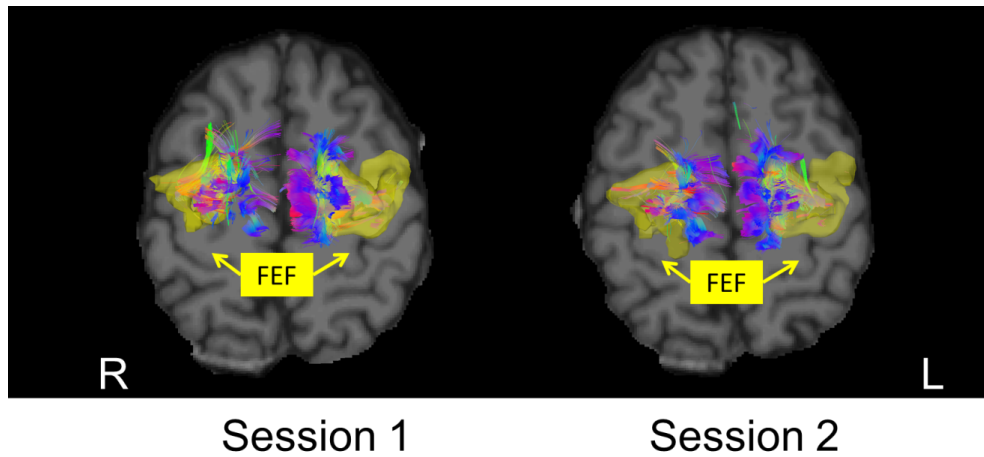


Figure 40. Fiber connections from FEF to the superior frontal gyrus. SFG projections represent probable connections to the supplementary eye fields (SEF). Top: axial view of representative tracts for participant 1. Images are presented in radiological orientation. Bottom: coronal view of SFG/SEF projections for participants 2–7. For these 6 participants, results are shown for only the first session; however, tractography and subsequent analysis were performed in all 14 available DSI datasets.

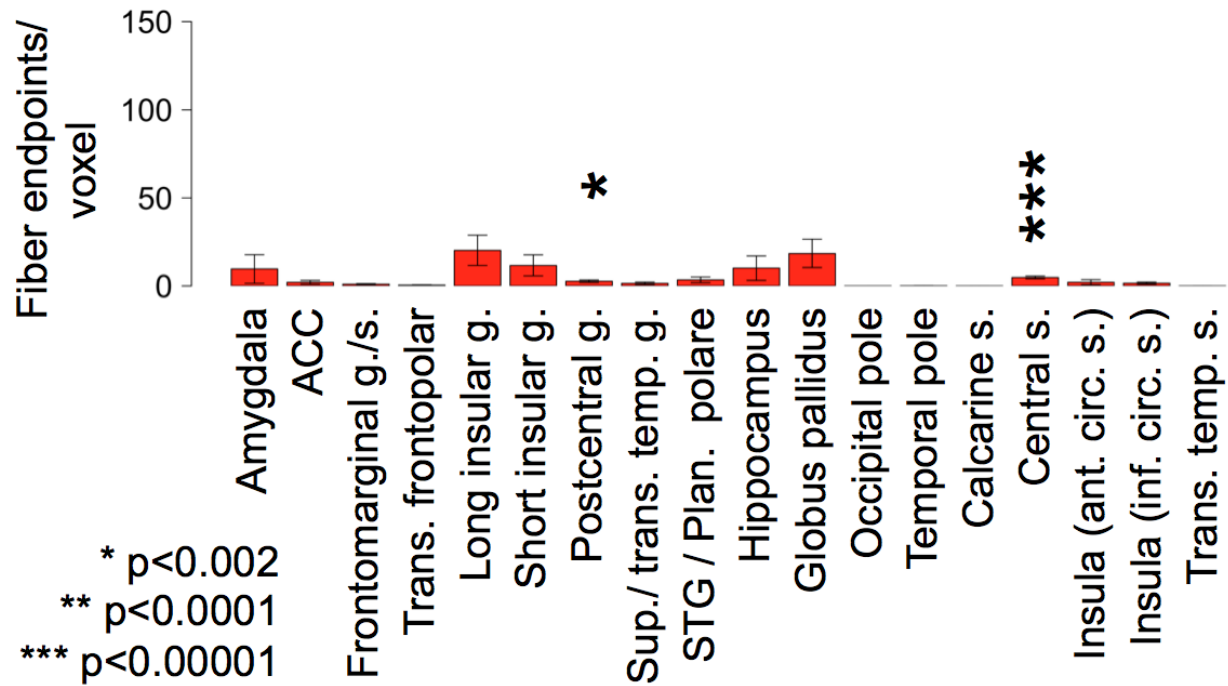


Figure 41. Mean fiber density of connections to foil ROIs. Asterisks indicate significant differences from a zero baseline, determined using a one-tailed t-test. Error bars represent 1 standard error above and below the mean. The abbreviation “g.” indicates gyrus; “s.” indicates sulcus. Fiber density is calculated as the total number of virtual fibers in a tract (and thus, the number of endpoints) divided by the number of voxels in the ROI.

Table 15. Significance of FEF connections to foil ROIs. Significance vs. a zero baseline is indicated by an asterisk; the threshold for significance was $p < 0.05/30$, representing a Bonferroni-adjusted familywise error rate of 0.05.

Region	t, df=27	p (uncorrected)	95% C.I. of connection density (fibers per voxel)	
Amygdala	1.2	0.3	-7.2	26.6
Anterior cingulate cortex	1.9	0.07	-0.1	4.1
Frontomarginal G/S	2.5	0.02	0.1	0.9
Transverse frontopolar G/S	2.5	0.02	0.0	0.3
Insula (lg/central S)	2.4	0.03	2.6	37.8
Short insular G	1.9	0.06	-0.7	24.2
*Postcentral G	3.7	0.001	1.2	4.2
STG/transverse temporal G	2.1	0.05	0.0	3.0
STG/polar plane	2.1	0.05	0.1	6.8
Hippocampus	1.4	0.2	-4.3	24.7
Pallidum	2.3	0.03	1.9	35.1
Occipital pole	0.0	1.0	0.0	0.0
Temporal pole	2.0	0.05	0.0	0.2
Calcarine S	1.0	0.3	0.0	0.1
*Central S	6.0	0.0000	3.1	6.4
Anterior insula (circular S)	1.5	0.15	-0.8	5.0
Inferior insula (circular S)	2.5	0.02	0.3	2.9
Transverse temporal S	1.4	0.2	0.0	0.1

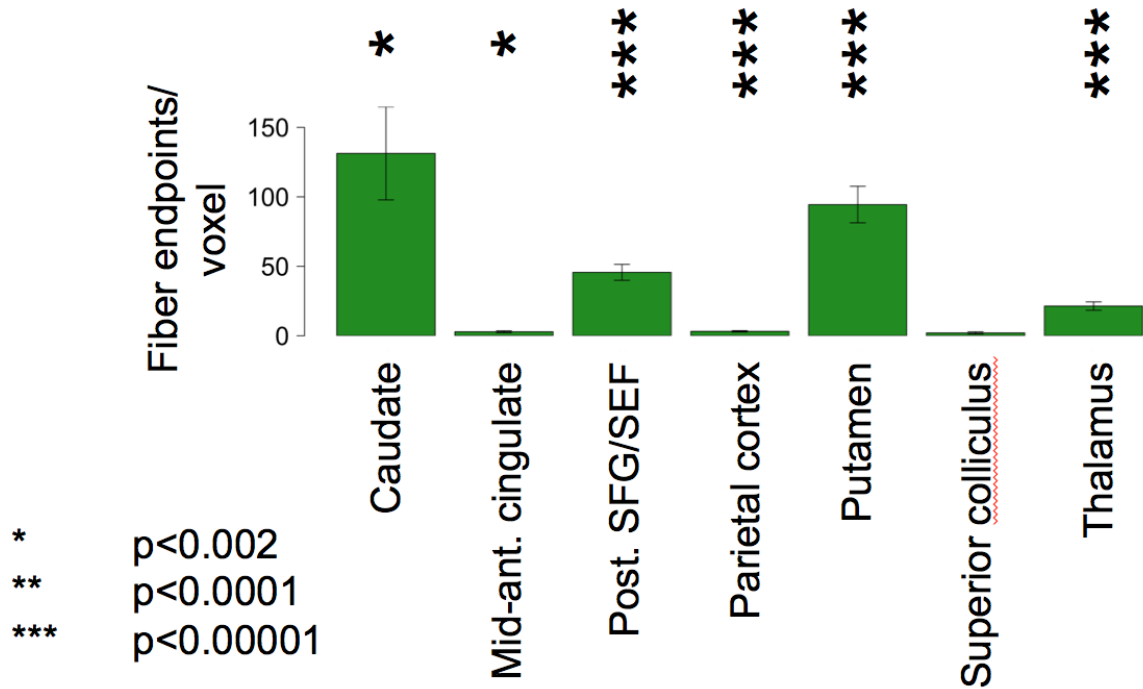


Figure 42. Mean fiber density of connections to target ROIs. Asterisks indicate significant differences from a zero baseline, determined using a one-tailed t-test. Error bars represent 1 standard error above and below the mean. Parietal cortex ROIs were collapsed into a single test. SFG/SEF indicates the superior frontal gyrus, including the supplementary eye field. Fiber density is calculated as the total number of virtual fibers in a tract (and thus, the number of endpoints) divided by the number of voxels in the ROI.

Table 16. Significance of FEF connections to target ROIs. Significance vs. a zero baseline is indicated by an asterisk; the threshold for significance was $p < 0.05/30$, representing a Bonferroni-adjusted familywise error rate of 0.05.

Region	t, df=27	p (uncorrected)	95% C.I. of connection density (fibers per voxel)	
*Caudate	3.9	0.0005	62.9	201.0
*Mid-anterior cingulate	4.3	0.0002	1.5	4.4
*Posterior SFG/SEF	8.0	0.0	34.0	57.6
*Angular gyrus	4.1	0.0003	3.6	10.8
Supramarginal gyrus	3.4	0.002	0.1	0.3
*Superior parietal lobe	3.7	0.0009	3.2	11.0
Precuneus	3.1	0.005	0.3	1.7
*Putamen	7.2	0.0	67.9	121.9
*IPS	3.9	0.0005	1.6	5.0
Subparietal sulcus	1.6	0.13	-0.1	0.5
Superior colliculus	2.5	0.02	0.4	3.7
*Thalamus	7.1	0.0000	15.3	27.6

Deterministic tractography results for target ROI connections are shown in Figures 38-40. Hypothesized projections from the FEF to target ROIs were found for nearly all participants and sessions (Table 17), although the density of these connections exhibited considerable variability. Post-hoc tests on the target set indicated significantly non-zero connections to 8 out of 12 ROIs, corresponding to a hit rate of 75 percent. The densest of these connections were to the posterior SFG (the anatomical location of the SEF), the thalamus, and the putamen. In some datasets, no FEF projections were found to the supramarginal gyrus, superior colliculus, or subparietal sulcus. The fiber density confidence intervals for these three ROIs and the precuneus had lower bounds near or below zero (Table 16); consequently, these connections were non-significant at the corrected significance threshold that we employed ($p = 0.05/30$, due to the 30 target and foil

ROIs tested per hemisphere). Connections to the supramarginal gyrus, in particular, closely missed significance at our adjusted criterion.

We next compared each connection from the target set to the total distribution of fiber density values in the foil set, using two-sample, two-tailed t-tests and assuming unequal variance between samples. The results confirmed that the caudate nucleus, posterior SFG/SEF, putamen, and thalamus exhibited significantly denser connectivity than did ROIs in the low-density foil set. However, connections to other target ROIs were not significantly denser than to foils. Of those ROIs that failed to show significantly non-zero connections, three (supramarginal gyrus, precuneus, and subparietal sulcus) had significantly lower mean fiber density than in the foil distribution.

The relative weakness of parietal connections, particularly to the IPS, was surprising given the functional significance of connections between FEF and parietal eyefields (Lynch & Tian, 2006). We speculated that reliable connections to the central sulcus and postcentral gyrus might have skewed the distribution of density values in the foil set and blurred its distinction from the target ROI set. However, the FEF-IPS connection remained statistically equivalent to the foil set after exclusion of the central sulcus and postcentral gyrus [$t(171)=-1.3$, $p<0.2$]; the same was found for the FEF-SPL connection [$t(48)=0.9$, $p<0.4$]. Under these test conditions, connections to the subparietal sulcus, supramarginal gyrus, and precuneus remained significantly sparser than the average connection to foil ROIs. Thus, the apparent sparseness of parietal projections did not result from the inclusion of outlier values in the low-density foil set. Rather, it may be due to the size of these areas: the anatomical ROIs employed in this analysis may include a large proportion of functionally non-relevant cortical surface, diluting the fiber density values that we employ.

Table 17. Count of detected FEF connections to target ROIs. Values indicate number of subjects, summed separately by hemisphere and session. Maximum possible count in each case is 7, based on the number of participants tested.

	Session 1		Session 2	
	Left	Right	Left	Right
Caudate	7	7	7	7
Mid-anterior cingulate	7	7	7	6
Posterior SFG/SEF	7	7	7	7
Angular gyrus	7	7	7	7
Supramarginal gyrus	6	5	6	6
Superior parietal lobe	7	7	7	7
Precuneus	7	7	7	5
Putamen	7	7	7	7
IPS	7	7	7	7
Subparietal sulcus	4	3	2	5
Superior colliculus	5	5	7	6
Thalamus	7	7	7	7

Table 18. Significance of target ROI connections relative to foil connection density. Baseline

was the mean density of FEF connections to foil ROIs. Significance is indicated by an asterisk; the threshold for significance was $p < 0.05/12$, representing a Bonferroni-adjusted familywise error rate of 0.05.

Region	t	df (adjusted)	p (uncorrected)
*Caudate	3.8	27	0.0008
Mid-anterior cingulate	-1.7	205	0.09
*Posterior SFG	7.0	29	0.0
Angular gyrus	1.1	46	0.3
*Supramarginal gyrus	-4.9	505	0.0
Superior parietal lobe	1.0	43	0.3
*Precuneus	-3.8	505	0.0002
*Putamen	6.8	27	0.0
IPS	-1.3	138	0.2
*Subparietal sulcus	-4.8	518	0.0
Superior colliculus	-2.3	147	0.03
*Thalamus	5.2	33	0.0

2.3.3 Topography of Target Projection Endpoints in FEF

Relatively little is known about the topography of fiber terminations within the FEF from various cortical and subcortical sources. To address this question, we plotted average endpoint fields (computed across all 14 datasets) for ROIs in the target set. The results are shown separately for subcortical and cortical sources in Figures 43 and 44. The densest and most widespread projections were from the putamen: these covered the fundus, anterior bank, and posterior bank of the precentral sulcus; they also extended anteriorly into the posterior portion of the superior frontal sulcus. Laterally, putamen projections reached approximately to the border between the upper and lower portions of the precentral sulcus. Medially, they included portions of the superior frontal gyrus, although a portion of these most medial endpoints may represent projections to the supplementary eye fields (SEF). Caudate projections were also widespread, and in the right hemisphere they extended further laterally than putamen fibers, to encompass part of the inferior portion of the precentral sulcus. Thalamic projections also extended laterally. Importantly, termination fields from all three subcortical sources appeared to have roughly the same center of mass, lying approximately in the precentral sulcus, near its intersection with the superior frontal sulcus and close to the middle frontal gyrus. This localization agrees with current consensus regarding the location of the human FEF (Amiez & Petrides, 2009; Paus, 1995; Rosano et al., 2002, 2003).

Cortical projections to the FEF were centered around the same location but exhibited more region-by-region heterogeneity in their termination fields (Figure 44). Projections from the SFG/SEF, like putamen projections, extended into the posterior portion of the SFS and covered the fundus and banks of the superior precentral sulcus. In contrast, projections from posterior

parietal cortex (PPC) were more selective: PPC terminations were densest in the SFS, just anterior to its junction with the precentral sulcus, and in the lateral portion of superior precentral sulcus. Middle/anterior cingulate cortex projections principally terminated at the dorsomedial extreme of the precentral sulcus, near the junction of the precentral and superior frontal gyri. These cingulate terminations were much more prominent in the left hemisphere than in the right.

2.3.4 Effects of Anisotropy Threshold

In addition to the results reported above, we tested FEF connectivity to target and foil ROIs at lower and higher anisotropy thresholds. In doing so, we sought to determine whether the observed pattern of results was specific to a single threshold, or else generalized across a range of possible parameter values. An ANOVA on tractography generated using a low QA threshold (half a standard deviation below the thresholds used above) revealed a main effect of ROI group [$F(1,830)=10.7$, $p<0.001$], but no main effect of hemisphere [$F(1,830)=0.0$, $p<1$] or interaction of ROI group with hemisphere [$F(1,830)=0.8$, $p<0.4$]. Tests of tractography generated at a high threshold indicated a qualitatively identical pattern of results: a significant effect of ROI group [$F(1,830)=56.6$, $p<0.0001$], but no effect of hemisphere [$F(1,830)=0.0$, $p<1$] or interactive effect [$F(1,830)=0.0$, $p<0.9$]. These results corroborated our initial findings, based on the anisotropy thresholds used in Experiment 1. Furthermore, they confirm that although fiber tracking results are influenced by tractography parameters, qualitative conclusions about fiber connectivity can remain consistent over substantial variability in procedural choices like tracking threshold.

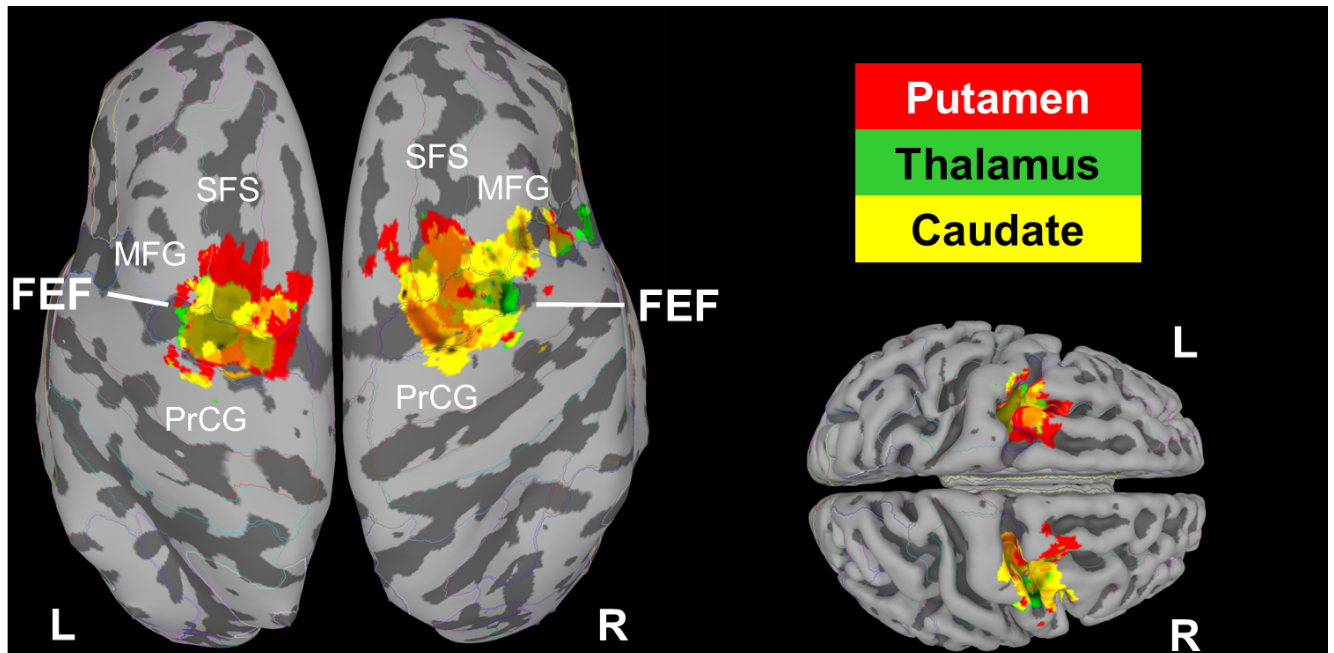


Figure 43. Endpoint fields of subcortical projections to FEF. Red=putamen, green=thalamus, yellow=caudate. Average results (across subjects and sessions) shown on an inflated surface representation of MNI template brain. Darker gray indicates concavities (i.e., sulci), while lighter gray indicates convexities (i.e., gyri). FEF label indicates frontal eye field location; MFG, middle frontal gyrus; SFS, superior frontal sulcus; PrCG, precentral gyrus. All maps were thresholded to mask out nodes with density less than 5 fibers.

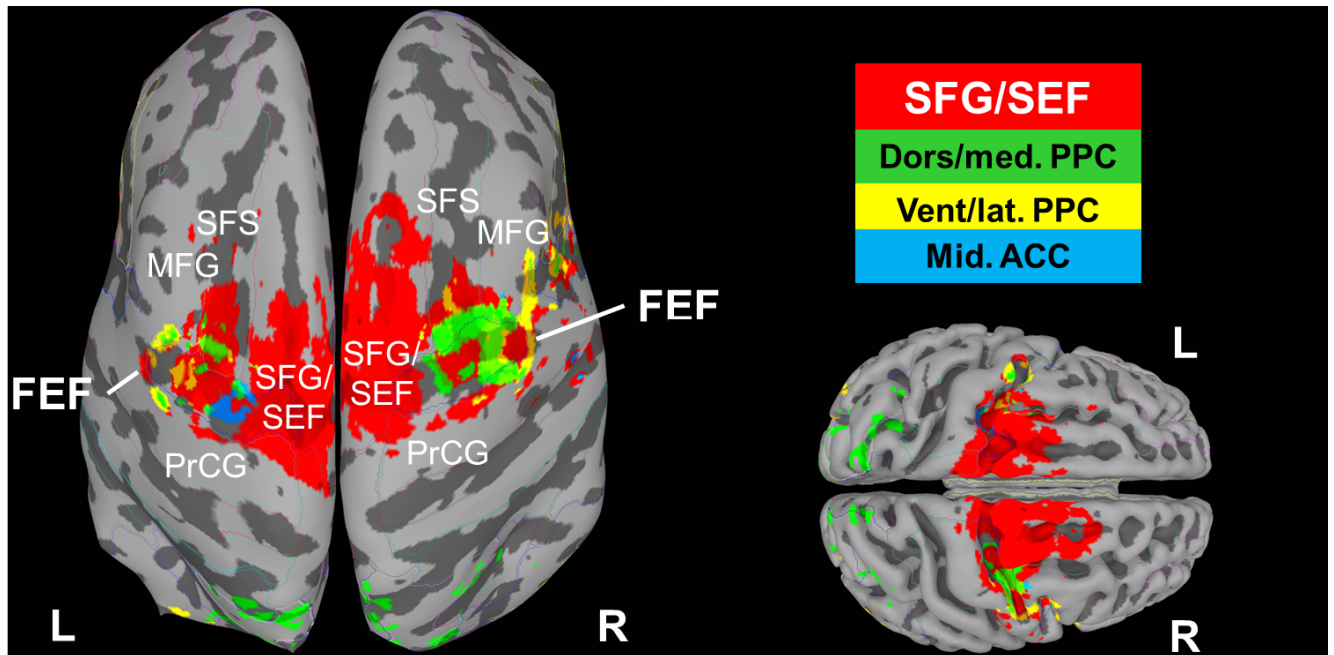


Figure 44. Endpoint fields of cortical projections to FEF. Red=superior frontal gyrus/supplementary eye field (SFG/SEF), green=dorsal and medial posterior parietal cortex (superior parietal lobule, precuneus, intraparietal sulcus, and subparietal sulcus), yellow=ventral and lateral posterior parietal cortex (angular and supramarginal gyri), blue=middle/anterior cingulate cortex. Average results (across subjects and sessions) shown on an inflated surface representation of MNI template brain. Darker gray indicates concavities (i.e., sulci), while lighter gray indicates convexities (i.e., gyri). FEF label indicates frontal eye field location; MFG, middle frontal gyrus; SFS, superior frontal sulcus; PrCG, precentral gyrus. All maps were thresholded to mask out nodes with density less than 5 fibers.

2.4 DISCUSSION

Tractography results broadly conformed to hypotheses based on extant knowledge of macaque brain connectivity. Expected FEF connections were consistent across sessions and individuals, particularly projections to the thalamus, striatum, and the supplementary eye fields. Importantly, connection density to subcortical targets and the SEF was significantly greater than in a set of foil connections, which we predicted to have sparser connections with the FEF.

Across datasets, the FEF was also reliably connected to the angular gyrus, superior parietal lobule (SPL), and intraparietal sulcus. We note that functional connectivity analysis has previously linked two of these regions (SPL and anterior IPS) to the FEF (Nelson et al., 2010). Despite their reliability, however, parietal connections were not reliably denser than connections to ROIs in the low-density foil set. Post-hoc testing suggested that this relative weakness, compared to low-density foil ROIs, was not a consequence of connections to the central sulcus and postcentral gyrus. Rather, we speculate that the wide fanning of FEF projections across parietal cortex may have diluted the density of these projections in any single parietal region.

We also detected reliable connections from the FEF to areas of the central sulcus and postcentral gyrus, which are associated with the posterior extent of primary motor cortex and with primary somatosensory cortex (Nieuwenhuys et al., p. 607, 623). In CoCoMac query results, corresponding reports in the macaque were limited to a single study, in which the FEF was not functionally identified (Barbas, 1988). Thus, it is unclear whether our findings indicate real connections with the FEF, or else false fiber streamlines resulting from confusion of FEF projections and independent inputs to motor and somatosensory cortex.

The variability of fiber density across datasets was not surprising, given our results in Experiment 1. We note that some foil ROIs showed dense connections in a few individual datasets (evident in the breadth of 95% confidence intervals in Table 15), but that those connections were ultimately non-significant when assessed across all participants, sessions, and hemispheres. Particularly dense projections were observed to the globus pallidus, hippocampus, and insula. Given the proximity of these areas to target ROIs, such as the thalamus, caudate, and putamen, we speculate that these dubious foil connections may result from false continuations or branchings off of valid FEF projection pathways. We note that while such false continuations represent a pervasive problem in DWI-based tractography, they are rarely reported or discussed.

We additionally found consistent hemispheric differences in the volume of the functionally-defined FEF; however, no hemispheric differences in FEF connection density or fiber count were found. It is unclear whether the differences in ROI mask size correspond to true differences in the spatial extent of the right vs. left FEF. Alternatively, the greater volume of right hemisphere FEF masks could be an artifact of the BOLD response: that is, more intense neuronal activation in FEF could lead to a larger metabolic response, distributed over a wider volume in the right hemisphere than the left. It is possible that more detailed analysis of the spatial dispersion of fiber endpoints in left vs. right FEF can help clarify this ambiguity.

2.5 INTERIM CONCLUSION

Comparison of DSI-based tractography to prior neuroanatomical knowledge represents one form of methodological validation. In the current study, hypothesis-driven investigation of the frontal eye field revealed known connections to both subcortical and cortical targets. We further compared the density of fiber projections to these targets with connections to an alternate set of brain areas hypothesized to have null or sparse connectivity with FEF. This comparison of targets and foils produced mixed results, suggesting possible overlap in the fiber density distributions for true and false connections. However, given the difficulty of making absolute predictions of connection presence or absence, it is also possible that some observed projections to foil set ROIs represent true but previously undocumented connections with the FEF.

Methods are still under development for applying fiber tractography to cognitive neuroscience hypotheses. Our research group has previously used DSI-based tractography to address questions about topographic relationships in fiber data, such as the preservation of somatotopic mapping in the brainstem (Verstynen et al., 2011) and the relation of attendotopic mapping in parietal cortex to retinotopic representations in early visual cortex (Greenberg et al., 2012). In contrast, the current paper presents an approach for testing hypotheses about relative connection density. Our results indicate that DSI-based tractography produces assessments of white matter connectivity which conform to extant data from invasive research in non-human animals. At the same time, these results confirm our findings in Experiment 1 of significant noise levels in tractography results, which can result from both random fluctuations in anisotropy estimates and systematic biases in tractography algorithms. The effect of this noise upon tract detection and reliability must be taken into account in applications of DSI-based tractography.

3.0 GENERAL DISCUSSION

3.1. RELIABILITY AND VALIDITY OF DSI TRACTOGRAPHY

The current study has made a significant contribution to the understanding of between-session reliability in DSI-based tractography, using complementary overlap and correlation metrics to illustrate both strengths and weaknesses of the method. We feel that these results present a valuable counterpoint to two previous studies on the reliability of DSI-based tractography (Bassett et al., 2011; Cammoun et al., 2012) , which reported near perfect replicability of global tracking results. In contrast, we have demonstrated the presence of significant noise levels within individual DSI datasets; fortunately, this noise appears to be remediable through between-session and between-subject comparison.

Based on results of the intercortical connectivity analysis (Part 1), we propose that averaging of multiple observations from a single individual is a viable (and perhaps preferable) alternative to post-hoc thresholding of a single dataset from that individual. Our analyses indicated that the most and least reliably detected fiber tracts spanned similar ranges of fiber density. This high degree of overlap suggests that thresholding tractography based on fiber density may both mask out valid fibers and fail to eliminate some false fibers. Of course, we note that Cammoun et al.

(2012) and Bassett et al. (2011) appear to have achieved high reliability scores partly through the use of thresholding methods based on fiber density or counts. It is possible that in these two studies, thresholding was sufficient to filter out most fibers resulting from randomly-distributed noise in the underlying diffusion data. However, the results of the current studies suggest that such thresholding might simultaneously have filtered out an unknown number of true but sparse fiber tracts.

Systematic examination of hit and false alarm rates in DWI-based tractography could help in determining the merits of post-hoc thresholding of brain graphs vs. other techniques for noise reduction. Our investigation of projections from the frontal eye fields (FEF) to various target and foil brain areas provided evidence of discriminability between true and false fiber distributions. In comparison to a zero baseline, eight of 12 target ROIs had significant FEF projections, while only 2 of 18 foil ROIs did. However, putative target connections had significant variability in fiber density, even for connections known to exist on the basis of evidence from invasive studies. Furthermore, only 4 of 12 target ROIs (caudate nucleus, posterior SFG/SEF, putamen, and thalamus) exhibited reliably greater fiber density than foil connections when compared directly. These findings again suggest that density- and count-based metrics may not be an optimal basis for removing noise from deterministic tractography results, due to the overlap in true and false fiber density distributions.

One caveat to the approach of reducing noise through multiple DSI scans is the variability that we observed in the intercortical connectivity analysis when all datasets were compared, collapsing across factors of subject and session. The set of MOD connections—that is, connections detected across all 14 DSI datasets—numbered only about 1000 of a possible 11175 intercortical connections and represented, on average, 24 % of the connections detected in any

single dataset. Further investigations should attempt to verify that this variability truly relates to brain connectivity differences between individuals. Alternatively, it is possible that similar variability would be observed by making a large number of observations within a single individual—for example, 14 DSI scans from the same participant, rather than the 2 scans performed in the current study. We note that prior studies by Bassett et al. (2011) and Cammoun et al. (2012) were also limited to only two observations per participant. If reliability scores reported by those authors and the current study are inflated by chance overlap in noisy fibers, including a larger number of replications per individual will reduce such overlap and yield a corrected reliability score.

3.2. PROCEDURAL CHOICES IN DETERMINISTIC TRACTOGRAPHY

The current study also examined a key procedural question that many DWI studies address in passing, if at all: namely, the choice between targeted and global approaches to seeding and fiber propagation. As we have demonstrated, these two approaches are characterized by large differences in detection power, as well as modest differences in between-session reliability. This difference is crucial to communicate to the broader imaging community, as more researchers begin to adopt DWI-based tractography into their research. At present, the global approach appears to be widespread in studies of whole-brain connectivity using deterministic tractography. The global approach is appealing because of its computational efficiency and because it ensures that all fibers of interest are generated using the same set of parameters. The converse of this

argument, of course, is that global seeding and propagation may compel the researcher to accept trade-offs, choosing tracking parameters and/or ROI constraints that are sub-optimal for some fibers of interest. Furthermore, global tractography usually involves a much larger array of possible seed points than targeted tractography, lowering the probability that any one trajectory will be sampled. Finally, a considerable downside of global tractography is the sheer volume of results generated: at present, the field has no automated tool for assessing the validity and quality of all the elements in a whole-brain connectivity matrix, even one based on simple anatomical units such as those provided by the FreeSurfer parcellation scheme. Given these practical concerns, we presently advocate for the use of targeted tractography to address limited anatomical hypotheses.

Of course, one concern with targeted tractography is that it may allow researchers, intentionally or unwittingly, to shape tractography results to match expectations. While this concern is reasonable, we feel that aspects of the current results help to mitigate the concern. In particular, a systematic parameter sensitivity test (Part 1) indicated that while tractography results did depend on user-set tractography variables, this dependency followed interpretable patterns, and accurate results were not limited to an arbitrary combination of parameter values. Likewise, our core finding in Part 2—greater fiber density in FEF projections to target than foil ROIs—was consistent over low, medium, and high values of the user-set tracking threshold (i.e., quantitative anisotropy).

On the basis of these findings, we conclude that tractography results are resistant to reasonable variations in key tractography variables. However, it is worthwhile to describe in greater detail the effects we observed in response to parameter manipulations. In the corticospinal tract parameter test of Part 1, we orthogonally manipulated tracking threshold,

angle threshold, and smoothing (a momentum parameter). Statistical analysis of accuracy scores for the resulting tracts suggested that angle and tracking thresholds exerted the largest effects upon results, both independently and in interaction with one another; these variables are also employed in other tractography software packages. In general, as one of these parameters became more lenient, it appeared that the other parameter must be more strictly controlled in order to maintain accuracy in the results. We found a range of optimal angle thresholds between 45° and 75° , spanning 3 of the 5 values tested; smaller values of this parameter (i.e., stricter angle thresholds) appeared too strict to produce valid fiber pathways. Interestingly, as the angle threshold was made laxer, accuracy declined if not accompanied by an increase in tracking threshold (Figure 27). With a stricter angle threshold, we found that a tracking threshold of $QA=0.5$ produced fairly accurate results; however, lower values appeared to reduce accuracy by allowing fiber propagation in areas of unstable diffusion contrast. In general, users should avoid very low tracking thresholds, although they may be useful in targeted tractography for detecting fibers with low diffusion contrast. In our study of FEF connections, for example, lowering the tracking threshold might have allowed us to detect projections to the superior colliculus in those datasets where none were visible at a threshold of $QA=0.5$. On the high end, QA values up to 1.0 allowed the generation of accurate tractography results, provided that they were coupled with lax angle thresholds.

We note that these specific values are unlikely to be relevant for most investigators, for two reasons. First, the corticospinal tract is among the structures with strongest diffusion contrast in the brain; while these fibers can still be detected at a threshold of $QA=1.0$, this threshold will probably be too strict for detecting sparse white matter structures or tracts that pass through areas of low diffusion contrast. Second, we note that the use of quantitative anisotropy as a basis for

thresholding diffusion data is peculiar to DSI Studio (at least presently). In contrast, most other tractography software packages use fractional anisotropy, which is not specific to a particular orientation and which scales differently from QA. Nevertheless, the pattern of interaction between angle and tracking thresholds should apply to other software packages and tracts of interest. One should thus perform empirical testing, preferably with targeted tracking, to determine the range of tracking thresholds best suited for detecting fibers of interest. This can be done by specifying a subset of data a priori for testing (should represent more than one individual, in order to ensure some generality across individuals), then applying similar values to other datasets.

One tractography parameter that we did not manipulate is step size, the linear distance that a virtual fiber grows every time the tractography algorithm chooses an orientation to proceed in. This variable is likely to interact heavily with angle threshold, since they jointly determine the curvature of a fiber—that is, at what angle does a fiber turn within some fixed distance. Small step sizes tend to allow the detection of more sharply curving fibers, while larger step sizes will tend to produce fibers with more gradual curves. It is also important to note that a short step size increases the computational intensity of tractography, as more segments have to be calculated in order to produce each fiber. A small step size may be useful for resolving fibers in complex configurations, such as crossings. However, researchers are likely to encounter a point of diminishing returns, as very small step sizes require sub-voxel spatial interpolation that is limited by the amount of information present in the acquired DSI data. Following the general trade-off principle described above, as step size becomes very small, angle threshold does not need to be so lenient; for example, a step size of 0.5 mm and an angle threshold of 0.70 is likely to be computationally inefficient, excessively lenient, and anatomically implausible. In the present

study, all tractography was performed with a step size of 1 mm; for comparison's sake, the length of a voxel edge during DSI acquisition was 3.5 mm.

3.3. ALTERNATIVE METRICS OF ANATOMICAL CONNECTION STRENGTH

Like most tractographic studies of brain connectivity, Experiments 1 and 2 used metrics of connectedness that are ultimately based on virtual fiber counts. Because virtual streamlines have no volume and are not modeled as having so by tractography programs, they are a less than ideal basis for quantifying connectivity. Furthermore, it is likely that fiber count-based metrics will not be comparable between different tractography algorithms and software packages, or across studies. Our own results in Experiment 1 indicate the arbitrariness of fiber count-based metrics, and their dependence on tracking conditions. In particular, targeted tracking produced much greater fiber density estimates and greater spatial coverage than global tracking. Furthermore, the reproducibility of fiber count-based metrics depended upon the context of the analysis: in our results, between-session correlation of fiber density values was high in ROI-based analysis, but was more variable and generally lower in voxelwise analysis. Thus, comparisons of fiber density values should be made with caution, ensuring that tractography results were generated under comparable tracking conditions. Likewise, comparison of statistics based on fiber density data (such as correlation coefficients) should be made with an awareness of the analysis context (e.g., spatial scale).

Ultimately, voxel-based anisotropy metrics may offer an alternative to connectivity measurements based on arbitrary fiber counts. In the current study, tracking threshold is set in terms of quantitative anisotropy (QA), a measurement of diffusion at a single orientation in three-dimensional space (Yeh et al., 2010). This directional specificity distinguishes QA from fractional anisotropy (FA), which is more commonly used as a basis for thresholding diffusion data. Both FA and QA are directionally non-specific: conceptually, they provide a summary measure of the “pointed-ness” or degree of constraint of diffusion within a voxel over all observed diffusion orientations, without indicating the specific orientation(s) of greatest diffusion (Jones, 2009). This generality makes FA unsuitable for estimating the strength of a particular anatomical connection.

In contrast, it may be possible to estimate connection strength by combining the QA values for ODF axes tangent to a fiber at each segment of its propagation. In this scenario, tractography would serve to label possible diffusion pathways; the number of fibers occupying each voxel in tractography results would be less relevant than characteristics of the QA values associated with a tract, such as the magnitude or variance of anisotropy, both along the tract’s length and over its cross-sectional area. However, devising a QA-based measurement of connection strength would require answering several challenging questions, such as whether to operate on individual fiber streamlines or on a reduced dataset, in which closely related streamlines were consolidated and/or fibers were clustered into discrete tracts. Relatedly, an anatomically plausible metric would have to treat QA as a finite quantity, which would rise or fall along the length of a tract as fibers converge into or diverge from it. Given these complexities, a QA-based measurement of connection strength represents a future objective, beyond the scope of the current study.

3.4. MERITS OF DWI-BASED TRACTOGRAPHY FOR LARGE-SCALE INVESTIGATION OF BRAIN CONNECTIVITY

Much recent research in cognitive neuroscience centers around integrating knowledge of brain structure with models of brain function. In human brain imaging studies, a detailed knowledge of brain connectivity can, at least in theory, allow researchers to constrain functional models, reducing uncertainty and increasing biological relevance. In studies of development and pathology, the ability to make repeated concurrent measurements of structural integrity and functional capacity may allow researchers to better understand the nature and timecourse of normal and abnormal changes in the brain. This contemporary interest in integrated structure/function studies is illustrated foremost in the NIH Human Connectome Project (<http://www.humanconnectome.org/>), which seeks to discover a template for human brain connectivity; such a template would probably supersede anatomical and cytoarchitectonic schema in providing a detailed, objective description of the average human brain.

It is interesting to speculate about what technologies and data formats will ultimately allow successful mapping of the human connectome. First and foremost, it is important that the biological basis of the derived information be well understood. Furthermore, because the connectivity of the human brain changes across the lifespan and as a result of pathology, a non-invasive technology that permits longitudinal studies is highly desirable; this will allow the collection of repeated measurements within specific individuals. Connectome mapping technology should have verified external validity, and any error or limits on its precision should be well quantified. For ease of analysis, connectome mapping results should be reported in a standardized digital format. Finally, these results should be scalable, so that they can be

examined at multiple levels of detail. This flexibility is important because it is not clear what the most meaningful or finest-grain units for describing the human connectome will turn out to be; in this respect, the connectome conceptually differs from its analogue, the human genome, which is ultimately reducible to an ordered list of DNA base pairs.

Despite its possible limitations in reliability and validity, we propose that DWI-based fiber tractography represents one of the most promising candidate technologies for large-scale, quantitative studies of brain connectivity. DWI-based fiber tractography can be used to examine white matter change or to look at changes in anatomical brain connectivity over the lifespan, observations which are otherwise impossible with invasive methods. It produces very high resolution data—fiber pathway information at millimeter or sub-millimeter scale—that can be easily scaled to coarser grain sizes. The essential physical cause of the diffusion signal is also clear—the constraint of water diffusion by white matter structures—even if important details are not, such as the quantitative relationship between the observed signal and actual diffusion volume (Basser & Özarslan, 2009). Until recently, the duration and expense of DSI data acquisition have made the collection of repeated observations impractical. However, optimized data acquisition is likely to substantially reduce this cost in the future, through innovations such as compressed sensing (Lee et al., 2012). Reduced time frames increase the feasibility of collecting multiple DSI scans, possibly within a single session, for clinical and research purposes alike. At present, the greatest shortcoming of DWI-based tractography for mapping brain connectivity remains the paucity of information about its reliability, external validity, noise levels, and effective spatial resolution. Although the current study partially addresses these questions, clarifying the methodological limits of DWI-based tractography will require further empirical and computational testing. These studies must elaborate upon the rates of true and false

fiber detection, the between-session and between-subject variability in tractography results, and the relationship between diffusion contrast and the real-world volume of diffusion in measured biological tissue.

4.0 CONCLUSION

The early adopters of a research technology must often deal with unresolved questions about proper procedure and statistical analysis; in confronting those questions, they tend to establish norms which later users will follow. Brain connectivity research using fiber tractography is presently in transition from a developmental stage to a period of wider usage by the cognitive neuroscience community, much as functional MRI was in the mid-1990s. Thus, standards and habits for performing tractography research are being rapidly established, and current users have a vested interest in clearly communicating both the capabilities and shortcomings of their methods to new users. We hope that the current study has succeeded in demonstrating DSI-based tractography's ability to reproduce known features of the brain's connective architecture, while also communicating the need for further validation and reliability testing. We note that the validation approach of Experiment 2, driven by brain connectivity findings from other species, represents one of multiple possible sources of validation. Alternative approaches include the use of hardware and software phantoms, which can demonstrate a technique's ability to resolve the ground truth in complex diffusion patterns; comparison to empirical results obtained with higher-resolution techniques, like polarized light imaging (Axer et al., 2011); and the comparison of abnormal tractography results to direct observations made during surgery (Fernandez-Miranda et

al., 2010) or to related behavioral deficits. With the current state of the art, we strongly advocate that investigators base their research on sufficient sample sizes, including multiple observations per individual, and that they rely on extant neuroanatomical findings from “gold-standard” sources in evaluating tractography results.

APPENDIX

EXPLANATION OF TRACTOGRAPHY VARIABLES

1. Fiber count: the number of virtual fibers which must be generated in order to complete tracking. Seed count and fiber count are mutually exclusive: only one can be set in a given tracking job. If fiber count is set, random seeds will continue to be generated until that limit is met. In practice, system memory limitations can cause tractography to fail before the fiber count is met, if a large enough number of seeds are generated.
2. Seed count: the number of random seeds generated in a single tractography job. While each of these seeds has the potential to generate a virtual fiber, in practice some proportion of these seeds fails for various reasons. Failure reasons include 1) stopping outside of allowed fiber lengths; 2) stopping outside of a user-set mask; 3) not passing through user-set ROIs. Seed count and fiber count are mutually exclusive: only one can be set in a given tracking job.
3. Smoothing: a momentum parameter to determine the direction of the next step in fiber propagation. Without smoothing, a fiber continues along the ODF axis closest in angle to its incoming direction—that is, the direction of propagation is independent of the fiber’s incoming direction. With smoothing, the incoming direction influences the direction of the next step, with the weight indicated by the smoothing value.

4. Step size: the length in mm of each propagation step. Larger step sizes are less computationally demanding and tend to produce smoother fiber trajectories; however, they may make the tractography algorithm less sensitive to sharp curvature in fibers. Tracking with smaller step sizes can increase computing time considerably and may produce “wavier” fibers.
5. Tracking threshold: a minimum anisotropy value for voxels. This threshold serves to mask out voxels as potential seed points. It also acts as a stopping criterion in tractography: when a fiber is propagated into a voxel and/or orientation below the threshold, propagation stops. When GQI reconstruction is used, this threshold represents quantitative anisotropy (QA), a measure specific to a single orientation and voxel. When DTI is used, the thresholded value is typically fractional anisotropy; other methods may use generalized fractional anisotropy.
6. Turning angle: a maximum angular deviation between steps during fiber propagation. If a fiber must turn at a sharper angle in order to propagate further, the fiber is terminated.

BIBLIOGRAPHY

- Amiez, C., & Petrides, M. (2009). Anatomical organization of the eye fields in the human and non-human primate frontal cortex. *Progress in Neurobiology*, 89(2), 220–230.
- Amos, A., 2000. A Computational Model of Information Processing in the Frontal Cortex and Basal Ganglia. *Journal of Cognitive Neuroscience*, 12(3), 505-519, MIT Press 2000.
- Axer, M., Amunts, K., Grässel, D., Palm, C., Dammers, J., Axer, H., Pietrzyk, U., et al. (2011). A novel approach to the human connectome: Ultra-high resolution mapping of fiber tracts in the brain. *NeuroImage*, 54(2), 1091–1101.
- Barbas, H. (1988). Anatomic organization of basoventral and mediodorsal visual recipient prefrontal regions in the rhesus monkey. *The Journal of Comparative Neurology*, 276(3), 313–342.
- Basser, P. J., & Özarslan, E. (2009). Introduction to Diffusion MR. *Diffusion MRI: From quantitative measurement to in-vivo neuroanatomy* (1st ed., pp. 3–10). London: Elsevier.
- Basser, P. J., & Pierpaoli, C. (2011). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance*, 213(2), 560–570.
- Basser, P. J., Mattiello, J., & LeBihan, D. (1994a). Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B*, 103(3), 247–254.
- Basser, P. J., Mattiello, J., & LeBihan, D. (1994b). MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1), 259–267.
- Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., & Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44(4), 625-632.
- Bassett, D. S., Brown, J. A., Deshpande, V., Carlson, J. M., Grafton, S. T. (2010). Conserved and variable architecture of human white matter connectivity. *NeuroImage*, 54:1262–79.

- Behrens, T. E. J., Johansen-Berg, H., Woolrich, M. W., Smith, S. M., Wheeler-Kingshott, C. A. M., Boulby, P. A., Barker, G. J., et al. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience*, 6(7), 750-757.
- Benjamini, Y., & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1), 289-300.
- Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences*, 1191(1), 133-155.
- Bernal, B., & Ardila, A. (2009). The role of the arcuate fasciculus in conduction aphasia. *Brain: A Journal of Neurology*, 132(9), 2309-2316.
- Bloom, J. S., & Hynd, G. W. (2005). The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition? *Neuropsychology Review*, 15(2), 59-71.
- Bürgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J. M., & Zilles, K. (2006). White matter fiber tracts of the human brain: Three-dimensional mapping at microscopic resolution, topography and intersubject variability. *NeuroImage*, 29(4), 1092-1105.
- Callaghan, P. T., MacGowan, D., Packer, K. J., Zelaya, F. O. (1990). High resolution q-space imaging in porous structures. *Journal of Magnetic Resonance*, 90, 177-182.
- Cammoun, L., Gigandet, X., Meskaldji, D., Thiran, J. P., Sporns, O., Do, K. Q., Maeder, P., et al. (2012). Mapping the human connectome at multiple scales with diffusion spectrum MRI. *Journal of Neuroscience Methods*, 203(2), 386-397.
- Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: History and current state. *Cortex*, 44(8), 953-961.
- Chang, L., Koay, C. G., Pierpaoli, C., & Basser, P. J. (2007). Variance of estimated DTI-derived parameters via first-order perturbation methods. *Magnetic Resonance in Medicine*, 57(1), 141-149.
- Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage*, 37(1), 343-360.
- Cole, M. W., Pathak, S., & Schneider, W. (2010). Identifying the brain's most globally connected regions. *NeuroImage*, 49(4), 3132-3148.
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3), 162-173.
- Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, 10(4-5), 171-178.

- Croxson, P. L., Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Pinski, M. A., Gross, C. G., Richter, W., Richter, M. C., Kastner, S., & Rushworth, M. F. S. (2005). Quantitative Investigation of Connections of the Prefrontal Cortex in the Human and Macaque Using Probabilistic Diffusion Tractography. *The Journal of Neuroscience*, 25(39), 8854–8866.
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31:968–80.
- Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., Fox, M. D., Snyder, A. Z., Vincent, J. L., Raichle, M. E., Schlaggar, B. L., & Petersen, S. E. (2007). Distinct Brain Networks for Adaptive and Stable Task Control in Humans. *Proceedings of the National Academy of Sciences*, 104(26), 11073–11078.
- Dum, R. P., & Strick, P. L. (2002). Motor areas in the frontal lobe of the primate. *Physiology & Behavior*, 77(4-5), 677–682.
- Farrell, J. A. D., Landman, B. A., Jones, C. K., Smith, S. A., Prince, J. L., van Zijl, P. C. M., & Mori, S. (2007). Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. *Journal of Magnetic Resonance Imaging: JMRI*, 26(3), 756–767.
- Fernandez-Miranda, J. C., Engh, J. A., Pathak, S. K., Madhok, R., Boada, F. E., Schneider, W., & Kassam, A. B. (2010). High-definition fiber tracking guidance for intraparenchymal endoscopic port surgery. *Journal of Neurosurgery*, 113(5), 990–999.
- Fernandez-Miranda, J. C., Pathak, S., Engh, J., Jarbo, K., Verstynen, T., Yeh, F.-C., Wang, Y., et al. (2012). High-definition fiber tractography of the human brain: neuroanatomical validation and neurosurgical applications. *Neurosurgery*, 71(2), 430–453.
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341–355.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., & Dale, A. M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1), 11–22.
- Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8(9), 700–711.
- Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: does the corpus callosum enable the human condition? *Brain: A Journal of Neurology*, 123(7), 1293–1326.

- Genovese, C. R., Noll, D. C., & Eddy, W. F. (1997). Estimating test-retest reliability in functional MR imaging I: Statistical methodology. *Magnetic Resonance in Medicine*, 38(3), 497-507.
- Goldman-Rakic, P. S. & Porrino, L. J. (1985). The primate mediodorsal (MD) nucleus and its projection to the frontal lobe. *The Journal of Comparative Neurology*, 242(4), 535–560.
- Goldman-Rakic, P.S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. *Handbook of Physiology: The Nervous System*, 5, 373-417.
- Gong, G., He, Y., Concha, L., Lebel, C., Gross, D. W., Evans, A. C., & Beaulieu, C. (2009). Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cerebral Cortex*, 19, 524–36.
- Greenberg, A. S., Verstynen, T., Chiu, Y.-C., Yantis, S., Schneider, W., & Behrmann, M. (2012). Visuotopic Cortical Connectivity Underlying Attention Revealed with White-Matter Tractography. *The Journal of Neuroscience*, 32(8), 2773–2782.
- Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *The Journal of Comparative Neurology*, 394(4), 475–495.
- Hagmann, P., Cammoun, L., Gigandet, X., Gerhard, S., Grant, E., Wedeen, V., Meuli, R., Thiran, J. P., Honey, C. J., Sporns, O. (2010). MR connectomics: principles and challenges. *Journal of Neuroscience Methods*, 194(1), 34–45.
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, 6, e159.
- Hagmann, P., Kuran, M., Gigandet, X., Thiran, P., Wedeen, V. J., Meuli, R., & Thiran, J.-P. (2007). Mapping human whole-brain structural networks with diffusion MRI. *PLoS One*, 2, e597.
- Hagmann P. From diffusion MRI to brain connectomics. Lausanne: Ecole Polytechnique Fédérale de Lausanne (EPFL), 2005.
- Hagmann, P., Gigandet, X., Meuli, R., Kötter, R., Sporns, O., & Wedeen, V. (2008). Quantitative Validation of MR Tractography Using the CoCoMac Database. *Proceedings of the International Society for Magnetic Resonance in Medicine*, 16.
- Harting, J. K., & Updyke, B. V. (2006). Oculomotor-related pathways of the basal ganglia. *Progress in Brain Research*, 151, 441–460.
- Heim, S., Hahn, K., Sämann, P. G., Fahrmeir, L., & Auer, D. P. (2004). Assessing DTI data quality using bootstrap analysis. *Magnetic Resonance in Medicine*, 52(3), 582-589.

- Hill, J., Inder, T., Neil, J., Dierker, D., Harwell, J., & Van Essen, D. (2010). Similar Patterns of Cortical Expansion During Human Development and Evolution. *Proceedings of the National Academy of Sciences*, 107(29), 13135–13140.
- Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., & Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6), 2035–2040.
- Hubbard, P. L., & Parker, G. J. M. (2009). Validation of Tractography. *Diffusion MRI: From quantitative measurement to in-vivo neuroanatomy* (1st ed., pp. 353–375). London: Elsevier.
- Huerta, M.F., Krubitzer, L.A., & Kaas, J.H. (1986). Frontal eye field as defined by intracortical microstimulation in squirrel monkeys, owl monkeys, and macaque monkeys: I. Subcortical connections. *The Journal of Comparative Neurology*, 253, 415–439.
- Huerta, M.F., Krubitzer, L.A., & Kaas, J.H. (1987). Frontal eye field as defined by intracortical microstimulation in squirrel monkeys, owl monkeys, and macaque monkeys. II. Cortical connections. *The Journal of Comparative Neurology*, 265, 332–361.
- Hutchison, R. M., Gallivan, J. P., Culham, J. C., Gati, J. S., Menon, R. S., & Everling, S. (2012). Functional connectivity of the frontal eye fields in humans and macaque monkeys investigated with resting-state fMRI. *Journal of Neurophysiology*, 107(9), 2463–2474.
- Jarbo, K., Verstynen, T., & Schneider, W. (2012). In vivo quantification of global connectivity in the human corpus callosum. *NeuroImage*, 59(3), 1988–1996.
- Jones, D. K. (2009). Gaussian Modeling of the Diffusion Signal. *Diffusion MRI: From quantitative measurement to in-vivo neuroanatomy* (1st ed., pp. 37–54). London: Elsevier.
- Killackey, H. P., Gould, H. J., 3rd, Cusick, C. G., Pons, T. P., & Kaas, J. H. (1983). The relation of corpus callosum connections to architectonic fields and body surface maps in sensorimotor cortex of new and old world monkeys. *The Journal of Comparative Neurology*, 219(4), 384–419.
- Klein, J. C., Behrens, T. E. J., Robson, M. D., Mackay, C. E., Higham, D. J., & Johansen-Berg, H. (2007). Connectivity-based parcellation of human cortex using diffusion MRI: Establishing reproducibility, validity and observer independence in BA 44/45 and SMA/pre-SMA. *NeuroImage*, 34(1), 204–211.
- Kötter R. (2004) Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics*, 2, 127-144.
- Lee, N., Wilkins, B., & Singh, M. (2012). Accelerated diffusion spectrum imaging via compressed sensing for the human connectome project. *Medical Imaging 2012: Image Processing*, 83144G.

- Leichnetz, G. R., Smith, D. J., & Spencer, R. F. (1984). Cortical projections to the paramedian tegmental and basilar pons in the monkey. *The Journal of Comparative Neurology*, 228(3), 388–408.
- Logan, B.R. and Rowe, D.B. (2004). An evaluation of thresholding techniques in fMRI analysis. *NeuroImage*, 22(1), 95-108.
- Lori, N. F., Akbudak, E., Shimony, J. S., Cull, T. S., Snyder, A. Z., Guillory, R. K., & Conturo, T. E. (2002). Diffusion tensor fiber tracking of human brain connectivity: acquisition methods, reliability analysis and biological results. *NMR in Biomedicine*, 15(7-8), 494-515.
- Lövdén, M., Bodammer, N. C., Kühn, S., Kaufmann, J., Schütze, H., Tempelmann, C., Heinze, H.-J., Düzel, E., Schmiedek, F., & Lindenberger, U. (2010). Experience-dependent plasticity of white-matter microstructure extends into old age. *Neuropsychologia*, 48(13), 3878–3883.
- Lynch, J. C., Graybiel, A. M., & Lobeck, L. J. (1985). The differential projection of two cytoarchitectonic subregions of the inferior parietal lobule of macaque upon the deep layers of the superior colliculus. *The Journal of Comparative Neurology*, 235(2), 241–254.
- Maitra, R., Roys, S. R., & Gullapalli, R. P. (2002). Test-retest reliability estimation of functional MRI data. *Magnetic Resonance in Medicine*, 48(1), 62-70.
- Makris, N., Papadimitriou, G. M., Sorg, S., Kennedy, D. N., Caviness, V. S., & Pandya, D. N. (2007). The occipitofrontal fascicle in humans: A quantitative, in vivo, DT-MRI study. *NeuroImage*, 37(4), 1100-1111.
- Martino, J., Brogna, C., Robles, S. G., Vergani, F., & Duffau, H. (2010). Anatomic dissection of the inferior fronto-occipital fasciculus revisited in the lights of brain stimulation data. *Cortex*, 46(5), 691–699.
- McFarland, N. R., & Haber, S. N. (2002). Thalamic Relay Nuclei of the Basal Ganglia Form Both Reciprocal and Nonreciprocal Cortical Connections, Linking Multiple Frontal Cortical Areas. *The Journal of Neuroscience*, 22(18), 8117–8132.
- Middleton F. A. and Strick, P.L, 2000. “Basal Ganglia Output and Cognition: Evidence from Anatomical, Behavioral, and Clinical Studies”, *Brain and Cognition*, 42, 183-200.
- Modha, D. S., & Singh, R. (2010). Network Architecture of the Long-Distance Pathways in the Macaque Brain. *Proceedings of the National Academy of Sciences*, 107(30), 13485–13490.
- Mori, S., Crain, B. J., Chacko, V. P., & Van Zijl, P. C. M. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45(2), 265–269.

- Narayanan, S. (2003). The role of cortico-basal-thalamic loops in cognition: a computational model and preliminary results. *Neurocomputing*, 52–54, 605–614.
- Nelson, S. M., Cohen, A. L., Power, J. D., Wig, G. S., Miezin, F. M., Wheeler, M. E., Velanova, K., Donaldson, D. I., Phillips, J. S., Schlaggar, B. L., & Petersen, S. E. (2010). A parcellation scheme for human left lateral parietal cortex. *Neuron*, 67(1), 156–170.
- Nieuwenhuys, R., Voogd, J., & van Huijzen, C. (2008). *The Human Central Nervous System* (4th ed.). Berlin: Springer.
- Noll, D. C., Genovese, C. R., Nystrom, L. E., Vazquez, A. L., Forman, S. D., Eddy, W. F., & Cohen, J. D. (1997). Estimating test-retest reliability in functional MR imaging II: Application to motor and cognitive activation studies. *Magnetic Resonance in Medicine*, 38(3), 508–517.
- Nomura, E. M., Gratton, C., Visser, R. M., Kayser, A., Perez, F., & D’Esposito, M. (2010). Double Dissociation of Two Cognitive Control Networks in Patients with Focal Brain Lesions. *Proceedings of the National Academy of Sciences*, 107(26), 12017–12022.
- Olszewski, J. *The Thalamus of the Macaca, Mulatta: An Atlas for Use with the Stereotaxic Instrument*. Basel: Karger, 1952.
- Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences*, 8(7), 315–324.
- Parthasarathy, H. B., Schall, J. D., & Graybiel, A. M. (1992). Distributed but Convergent Ordering of Corticostriatal Projections: Analysis of the Frontal Eye Field and the Supplementary Eye Field in the Macaque Monkey. *The Journal of Neuroscience*, 12(11), 4468–4488.
- Paus, T. (1996). Location and function of the human frontal eye-field: A selective review. *Neuropsychologia*, 34(6), 475–483.
- Pfefferbaum, A., Adalsteinsson, E., & Sullivan, E. V. (2003). Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. *Journal of Magnetic Resonance Imaging: JMRI*, 18(4), 427–433.
- Phillips, J. S., Greenberg, A. S., Pyles, J. A., Pathak, S. K., Behrmann, M., Schneider, W., & Tarr, M. J. (in press). Co-Analysis of Brain Structure and Function using fMRI and Diffusion-Weighted Imaging. *Journal of Visualized Experiments*.
- Pyles, J. A., Verstynen, T. D., Schneider, W., & Tarr, M. J. (2011). Structural connectivity of face selective cortical regions with high definition fiber tracking. *Journal of Vision*, 11(11), 655.
- Rogers, B. P., Morgan, V. L., Newton, A. T., & Gore, J. C. (2007). Assessing functional connectivity in the human brain by fMRI. *Magnetic Resonance Imaging*, 25(10), 1347–1357.

- Romanski, L. M., Giguere, M., Bates, J. F., & Goldman-Rakic, P. S. (1997). Topographic organization of medial pulvinar connections with the prefrontal cortex in the rhesus monkey. *The Journal of Comparative Neurology*, 379(3), 313–332.
- Rosano, C., Krisky, C. M., Welling, J. S., Eddy, W. F., Luna, B., Thulborn, K. R., & Sweeney, J. A. (2002). Pursuit and saccadic eye movement subregions in human frontal eye field: a high-resolution fMRI investigation. *Cerebral Cortex*, 12(2), 107–115.
- Rosano, C., Sweeney, J. A., Melchitzky, D. S., & Lewis, D. A. (2003). The human precentral sulcus: chemoarchitecture of a region corresponding to the frontal eye fields. *Brain Research*, 972(1-2), 16–30.
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52(3), 1059–1069.
- Sakaie, K. E., & Lowe, M. J. (2007). An objective method for regularization of fiber orientation distributions derived from diffusion-weighted MRI. *NeuroImage*, 34(1), 169–176.
- Schmahmann & Pandya (2007). The complex history of the fronto-occipital fasciculus.
- Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D’Arceuil, H. E., de Crespigny, A. J., & Wedeen, V. J. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain: A Journal of Neurology*, 130(3), 630–653.
- Seltzer, B., & Pandya, D. N. (1984). Further observations on parieto-temporal connections in the rhesus monkey. *Experimental Brain Research*, 55(2), 301–312.
- Seunarine, K. K., & Alexander, D. C. (2009). Multiple Fibers: Beyond the Diffusion Tensor. *Diffusion MRI: From quantitative measurement to in-vivo neuroanatomy* (1st ed.). London: Elsevier.
- Shook, B. L., Schlag-Rey, M., & Schlag, J. (1991). Primate supplementary eye field. II. Comparative aspects of connections with the thalamus, corpus striatum, and related forebrain nuclei. *The Journal of Comparative Neurology*, 307(4), 562–583.
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Yongyue, De Stefano, N, Brady, M. J., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(S1), S208–S219.
- Stanton, G. B., Deng, S-Y., Goldberg, E. M., & McMullen, N. T. (1989). Cytoarchitectural characteristic of the frontal eye fields in macaque monkeys. *The Journal of Comparative Neurology*, 282(3), 415–427.

- Stanton, G.B., Goldberg, M.E. and Bruce, C.J. (1988). Frontal eye field efferents in the macaque monkey: II. Topography of terminal fields in midbrain and pons. *The Journal of Comparative Neurology*, 271, 493–506.
- Stephan, K. E., Kamper, L., Bozkurt, A., Burns, G. A., Young, M. P. & Kotter, R. (2001). Advanced database methodology for the collation of connectivity data on the Macaque brain (CoCoMac). *Philosophical Transactions of the Royal Society, Series B, Biological Sciences*, 356, 1159–1186.
- Thottakara, P., Lazar, M., Johnson, S. C., & Alexander, A. L. (2006). Application of Brodmann's area templates for ROI selection in white matter tractography studies. *NeuroImage*, 29, 868–78.
- Tian, J. R. & Lynch, J. C. (1996). Functionally defined smooth and saccadic eye movement subregions in the frontal eye field of Cebus monkeys. *Journal of Neurophysiology*, 76(4), 2740–2753.
- Tomassini, V., Jbabdi, S., Klein, J. C., Behrens, T. E. J., Pozzilli, C., Matthews, P. M., Rushworth, M. F. S., et al. (2007). Diffusion-weighted imaging tractography-based parcellation of the human lateral premotor cortex identifies dorsal and ventral subregions with anatomical and functional specializations. *The Journal of Neuroscience*, 27(38), 10259–10269.
- Tournier, J.-D., Calamante, F., Gadian, D. G., & Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23(3), 1176–1185.
- Tournier, J., Mori, S., & Leemans, A. (2011). Diffusion tensor imaging and beyond. *Magnetic Resonance in Medicine*, 65(6), 1532–1556.
- Tuch, D. S. (2004). Q-ball imaging. *Magnetic Resonance in Medicine*, 52(6), 1358–1372.
- Tuch, D. S., Wisco, J. J., Khachaturian, M. H., Ekstrom, L. B., Kötter, R., & Vanduffel, W. (2005). Q-ball imaging of macaque white matter architecture. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 360(1457), 869–879.
- Türe, U., Yasargil, M. G., & Pait, T. G. (1997). Is there a superior occipitofrontal fasciculus? A microsurgical anatomic study. *Neurosurgery*, 40(6), 1226–1232.
- Verstynen, T., Jarbo, K., Pathak, S., & Schneider, W. (2011). In Vivo Mapping of Microstructural Somatotopies in the Human Corticospinal Pathways. *Journal of Neurophysiology*, 105(1), 336–346.
- Walker, A. E. (1940). A cytoarchitectural study of the prefrontal area of the macaque monkey. *The Journal of Comparative Neurology*, 73(1), 59–86.
- Wedeen, V. J., Wang, R. P., Schmahmann, J. D., Benner, T., Tseng, W. Y. I., Dai, G., Pandya, D. N., Hagmann, P., D'Arceuil, H., & de Crespigny, A. J. (2008). Diffusion spectrum

- magnetic resonance imaging (DSI) tractography of crossing fibers. *NeuroImage*, 41(4), 1267-1277.
- Wedeen, V. J., Hagmann, P., Tseng, W. I., Reese, T. G., & Weisskoff, R. M. (2005). Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine*, 54(6), 1377-1386.
- Westlye, L. T., Walhovd, K. B., Dale, A. M., Bjørnerud, A., Due-Tønnessen, P., Engvig, A., Grydeland, H., Tamnes, C. K., Østby, Y., Fjell, A. M. (2010). Life-Span Changes of the Human Brain White Matter: Diffusion Tensor Imaging (DTI) and Volumetry. *Cerebral Cortex*, 20(9), 2055–2068.
- Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45(S1), S173-S186.
- Yeh, F. C., Wedeen, V. J., & Tseng, W.-Y. I. (2010). Generalized Q-Sampling Imaging. *IEEE Transactions on Medical Imaging*, 29(9), 1626-1635.
- Zou, K. H., Greve, D. N., Wang, M., Pieper, S. D., Warfield, S. K., White, N. S., Manandhar, S., Brown, G. G., Vangel, M. G., Kikinis, R., & Wells, W. M. (2005). Reproducibility of Functional MR Imaging: Preliminary Results of Prospective Multi-institutional Study Performed by Biomedical Informatics Research Network. *Radiology*, 237(3), 781 -789.